

657 '05
AC 57-R

THE ACCOUNTING REVIEW

VOLUME 77, No. 1 JANUARY 2002

PETER O. CHRISTENSEN, GERALD A. FELTHAM, and MARTIN G. H. WU
"Cost of Capital" in Residual Income for Performance Evaluation

STEPHEN P. BAGINSKI, JOHN M. HASSELL, and MICHAEL D. KIMBROUGH
The Effect of Legal Environment on Voluntary Disclosure: Evidence from Management Earnings Forecasts Issued in U.S. and Canadian Markets

T. JEFFREY WILKS
Predecisional Distortion of Evidence as a Consequence of Real-Time Audit Review

JEFF P. BOONE
Revisiting the Reportedly Weak Value Relevance of Oil and Gas Asset Present Values: The Roles of Measurement Error, Model Misspecification, and Time-Period Idiosyncrasy

HOLLIS ASHBAUGH and PER OLSSON
An Exploratory Study of the Valuation Properties of Cross-Listed Firms' IAS and U.S. GAAP Earnings and Book Values

MORTON PINCUS and SHIVARAM RAJGOPAL
The Interaction between Accrual Management and Hedging: Evidence from Oil and Gas Firms

MICHAEL J. SMITH
Ex Ante and Ex Post Discretion over Arm's Length Transfer Prices

D. JORDAN LOWE, PHILIP M. J. RECKERS, and STACEY M. WHITCOTTON
The Effects of Decision-Aid Use and Reliability on Jurors' Evaluations of Auditor Liability

MICHAEL R. KINNEY and WILLIAM F. WEMPE
Further Evidence on the Extent and Origins of JIT's Profitability Effects

AMERICAN ACCOUNTING ASSOCIATION

The members of the 2001–2002 Executive Committee of the Association are as follows:

President	Joel S. Demski, University of Florida
President-Elect	G. Peter Wilson, Boston College
Vice President	William E. (Bill) McCarthy, Michigan State University
Vice President–Education	Kevin D. Stocks, Brigham Young University
Vice President–Finance	Larry E. Rittenberg, University of Wisconsin–Madison
Vice President–International	Judy S. L. Tsui, City University of Hong Kong
Vice President–Professional Relations	Samuel A. Vitkoski, BDO Seidman
Vice President–Publications	Lawrence A. Tomassini, The Ohio State University
Vice President–Research	Ronald A. Dye, Northwestern University
Vice President–Sections and Regions	Alexander L. Gabbin, James Madison University

The full-time Executive Director of the Association is Craig E. Polhemus.

The American Accounting Association was founded in 1916 as the American Association of University Instructors in Accounting. Its present name was adopted in 1936. The Association is a voluntary organization of persons interested in accounting education, research, and practice. Full membership entitles one to receive *Accounting Education News*, a selection of one, two, or all three of the Association-wide journals (*The Accounting Review*, *Accounting Horizons*, and *Issues in Accounting Education*) and the opportunity to purchase Association publications at a discount.

Membership in the Association is available at the following annual rates:

Full membership:

The membership dues are based on the selection of one, two, or all three of the following Association-wide journals: *The Accounting Review*, *Accounting Horizons*, and *Issues in Accounting Education*.

(This does not include Section journals.)

Select one Association-wide journal	U.S. \$85.00
Select two Association-wide journals	U.S. \$95.00
All three Association-wide journals	U.S. \$105.00

Associate membership for students residing in the United States and Canada:

Select one Association-wide journal	U.S. \$20.00
Select two Association-wide journals	U.S. \$30.00
All three Association-wide journals	U.S. \$40.00

Electronic journal option available for an additional U.S. \$20.00 to above-mentioned full-membership and associate-membership dues.

In addition, full members may belong to one or more Sections: Accounting, Behavior and Organizations; Accounting Programs Leadership Group; American Taxation Association; Artificial Intelligence/Emerging Technologies; Auditing; Financial Accounting and Reporting; Gender Issues in Accounting; Government and Nonprofit; Information Systems; International Accounting; Management Accounting; Public Interest; Teaching and Curriculum; and Two-Year College. Each Section assesses its own annual dues, which are collected by the Association.

Editorial correspondence and all manuscripts should be sent to Professor Linda Smith Bamber, Editor, *The Accounting Review*, Tull School of Accounting, Terry College of Business, University of Georgia, 255 Brooks Hall, Athens, GA 30602-6252.

The submission fee of \$75.00 for members of the AAA or \$100.00 for nonmembers must accompany manuscripts when submitted. Books for review should be sent to Jane L. Reimers, College of Business, Florida State University, Tallahassee, FL 32306-1042. Request for reprint permission, subscriptions, advertisements, and dues should be sent to American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399 • Telephone: (941) 921-7747 • Fax: (941) 923-4093 • Email: Office@aaahq.org. Notice of change of address should be mailed at least eight weeks in advance to ensure undelayed receipt of *The Accounting Review*.

The Accounting Review (ISSN 0001-4826) is published four times a year in January, April, July, and October by the American Accounting Association, 5717 Bessie Drive, Sarasota, Florida 34233-2399. General subscriptions are available at the annual rate of U.S. \$160; membership subscription rate is \$25 with discounts available; and emeritus subscription rate is \$8.75. Periodical postage paid at Sarasota, Florida and additional mailing offices. **POSTMASTER:** Send address change to ***The Accounting Review***, 5717 Bessie Drive, Sarasota, Florida 34233-2399.

R, 10, 644

THE ACCOUNTING REVIEW

A QUARTERLY JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

Editor

Linda Smith Bamber

Consulting Editor

E. Michael Bamber

Associate Editors

Greg Clinch

Raffi J. Indjejikian

Steven J. Kachelmeier

Kenneth J. Klassen

Marlys Gascho Lipe

Joan L. Luft

Mark W. Nelson

Grace Pownall

Richard C. Sansing

Daniel B. Thornton

Senyo Tse

27

VOL. 77

JANUARY 2002

No. 1

PETER O. CHRISTENSEN, GERALD A. FELTHAM, and MARTIN G. H. WU	
"Cost of Capital" in Residual Income for Performance Evaluation	1
STEPHEN P. BAGINSKI, JOHN M. HASSELL, and MICHAEL D. KIMBROUGH	
The Effect of Legal Environment on Voluntary Disclosure: Evidence from Management	
Earnings Forecasts Issued in U.S. and Canadian Markets	25
T. JEFFREY WILKS	
Predecisional Distortion of Evidence as a Consequence of Real-Time Audit Review	51
JEFF P. BOONE	
Revisiting the Reportedly Weak Value Relevance of Oil and Gas Asset Present Values: The	
Roles of Measurement Error, Model Misspecification, and Time-Period Idiosyncrasy	73
HOLLIS ASHBAUGH and PER OLSSON	
An Exploratory Study of the Valuation Properties of Cross-Listed Firms' IAS and U.S. GAAP	
Earnings and Book Values	107
MORTON PINCUS and SHIVARAM RAJGOPAL	
The Interaction between Accrual Management and Hedging: Evidence from Oil and Gas Firms	
.....	127
MICHAEL J. SMITH	
Ex Ante and Ex Post Discretion over Arm's Length Transfer Prices	161
D. JORDAN LOWE, PHILIP M. J. RECKERS, and STACEY M. WHITCOTTON	
The Effects of Decision-Aid Use and Reliability on Jurors' Evaluations of Auditor Liability	185
MICHAEL R. KINNEY and WILLIAM F. WEMPE	
Further Evidence on the Extent and Origins of JIT's Profitability Effects	203

BOOK REVIEWS, Jane L. Reimers, Editor

PricewaterhouseCoopers LLP and SAP AG, <i>The E-Business Workplace: Discovering the Power of</i>	
<i>Enterprise Portals</i>	WILLIAM A. HILLISON 227
Young, S. David, and Stephen F. O'Byrne, <i>EVA® and Value-Based Management: A Practical Guide to</i>	
<i>Implementation</i>	JENNIFER FRANCIS 228
Editorial Data	229
Occasional Reviewers and Associate Editors	230
Editorial Policy and Style Information	233

Editorial Staff of *The Accounting Review*
2001–2002

EDITOR

Linda Smith Bamber, University of Georgia

CONSULTING EDITOR

E. Michael Bamber, University of Georgia

ASSOCIATE EDITORS

Greg Clinch, Melbourne Business School

Raffi J. Indjejikian, University of Michigan

Steven J. Kachelmeier, The University of Texas at Austin

Kenneth J. Klassen, University of Waterloo

Marlys Gascho Lipe, University of Oklahoma

Joan L. Luft, Michigan State University

Mark W. Nelson, Cornell University

Grace Pownall, Emory University

Richard C. Sansing, Dartmouth College

Daniel B. Thornton, Queen's University

Senyo Tse, The University of Texas at Austin

BOOK REVIEW EDITOR

Jane L. Reimers, Florida State University

American Accounting Association

EXECUTIVE DIRECTOR

Craig E. Polhemus

COMMUNICATIONS DEPARTMENT

Beverly Harrelson

Carol Stanton

Diane Hazard

Janet Wood

Kathy Casper

Jan Kovarik (subcontracted)

Helen Griffith

OTHER AAA STAFF

Tracey E. Sutherland

Judy Cothern

Debora Gardner

James Szelminski

Roland LaTulip

Linda Watterworth

Patricia A. Hoban

Diana Lombardi

Deirdre Harris

Barbara Brady

Mary Cole

Editorial Advisory and Review Board

Ashiq Ali, The University of Arizona

Shannon W. Anderson, Rice University

Benjamin C. Ayers, University of Georgia

Frances L. Ayres, University of Oklahoma

William R. Baber, The George Washington University

Stephen P. Baginski, Indiana University

Ramamurthy Balakrishnan, The University of Iowa

Orie E. Barron, Pennsylvania State University

Jan Barton, Emory University

Anne L. Beatty, Pennsylvania State University

Jacob G. Birnberg, University of Pittsburgh

Robert J. Bloomfield, Cornell University

Robert M. Bowen, University of Washington

Lawrence D. Brown, Georgia State University

David C. Burgstahler, University of Washington

Brian J. Bushee, University of Pennsylvania

Robert M. Bushman, The University of North Carolina
at Chapel Hill

Carolyn M. Callahan, University of Arkansas

Youngsoon Susan Cheon, Chung-Ang University

Kevin C. W. Chen, Hong Kong University of Science &
Technology

Chee W. Chow, San Diego State University

C. Bryan Cloyd, University of Illinois at Urbana-
Champaign

Paul A. Copley, University of Georgia

Patricia M. Dechow, University of Michigan

Mark L. DeFond, University of Southern California

Ilya D. Dichev, University of Michigan

Sunil Dutta, University of California, Berkeley

Pieter T. Elgers, University of Massachusetts

Merle M. Erickson, University of Chicago

John (Harry) Evans III, University of Pittsburgh

Paul E. Fischer, Pennsylvania State University

Joseph G. Fisher, Indiana University

Jere R. Francis, University of Missouri–Columbia

O. Finley Graves, Kansas State University

Wayne R. Guay, University of Pennsylvania

David A. Guenther, University of Colorado at Boulder

Mahendra Gupta, Washington University

Susan F. Haka, Michigan State University

John R. Hand, The University of North Carolina at
Chapel Hill

Donald R. Herrman, Oregon State University

D. Eric Hirst, The University of Texas at Austin

Chris E. Hogan, Southern Methodist University

Patrick E. Hopkins, Indiana University

K. E. (Skip) Hughes II, Louisiana State University

Christopher D. Ittner, University of Pennsylvania

Debra C. Jeter, Vanderbilt University

James J. Jiambalvo, University of Washington

Deen Kemsley, Columbia University

S. Jane Kennedy, University of Washington

William R. Kinney, Jr., The University of Texas at
Austin

Lisa L. Koonce, The University of Texas at Austin

William N. Lanen, University of Michigan

Charles M. C. Lee, Cornell University

Robert Libby, Cornell University

Robert C. Lipe, University of Oklahoma

Russell J. Lundholm, University of Michigan

Joseph Magliolo III, Southern Methodist University

Laureen A. Maines, Indiana University
Ella Mae Matsumura, University of Wisconsin-Madison
Steven R. Matsunaga, University of Oregon
Linda S. McDaniel, The University of North Carolina at Chapel Hill
Maureen F. McNichols, Stanford University
Richard R. Mendenhall, University of Notre Dame
Lillian F. Mills, The University of Arizona
James N. Myers, University of Illinois at Urbana-Champaign
Patricia C. O'Brien, University of Waterloo
Mark E. Peecher, University of Illinois at Urbana-Champaign
Kathy R. Petroni, Michigan State University
Jamie Pratt, Indiana University
Robert J. Ramsay, University of Kentucky
Madhav V. Rajan, University of Pennsylvania
Stephen G. Ryan, New York University
Katherine Schipper, Financial Accounting Standards Board

D. Shores, University of Washington
Paul J. Simko, Emory University
Richard G. Sloan, University of Michigan
J. Reed Smith, SUNY at Buffalo
Geoffrey B. Sprinkle, Indiana University
Thomas L. Stober, University of Notre Dame
Steve G. Sutton, University of Connecticut/University of Melbourne
Hun-Tong Tan, Nanyang Technological University
Siew Hong Teoh, The Ohio State University
Lynda Thoman, Purdue University
Richard Tubbs, The University of Iowa
Igor Vaysman, INSEAD
Robert E. Verrecchia, University of Pennsylvania
Martha L. Wartick, University of Northern Iowa
Gregory B. Waymire, Emory University
Hong Xie, University of Illinois at Urbana-Champaign
Teri L. Yohn, Georgetown University

PAST EDITORS

1926–1929	William A. Paton	1971–1972	Eldon S. Hendriksen
1929–1943	Eric L. Kohler	1973–1975	Thomas F. Keller
1944–1947	A. C. Littleton	1976–1978	Don T. DeCoster
1948–1949	Robert L. Dixon	1978–1983	Stephen A. Zeff
1950–1959	Frank P. Smith	1983–1987	Gary L. Sundem
1960–1962	Robert K. Mautz	1987–1990	William R. Kinney, Jr.
1963–1964	Lawrence L. Vance	1990–1994	A. Rashad Abdel-khalik
1965–1967	Wendell Trumbull	1994–1997	Robert P. Magee
1968–1970	Charles H. Griffin	1997–2000	Gerald L. Salamon

THE ACCOUNTING REVIEW

A QUARTERLY JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

Editor

Linda Smith Bamber

Consulting Editor

E. Michael Bamber

Associate Editors

Greg Clinch

Raffi J. Indjejikian

Steven J. Kachelmeier

Kenneth J. Klassen

Marlys Gascho Lipe

Joan L. Luft

Mark W. Nelson

Grace Pownall

Richard C. Sansing

Daniel B. Thornton

Senyo Tse

VOL. 77

APRIL 2002

No. 2

STEPHEN H. PENMAN and XIAO-JUN ZHANG

Accounting Conservatism, the Quality of Earnings, and Stock Returns 237

RONALD R. KING

An Experimental Investigation of Self-Serving Biases in an Auditing Trust Game: The Effect of Group Affiliation 265

ROBERT M. BOWEN, ANGELA K. DAVIS, and DAWN A. MATSUMOTO

Do Conference Calls Affect Analysts' Forecasts? 285

CRISTI A. GLEASON and LILLIAN F. MILLS

Materiality and Contingent Tax Liability Reporting 317

THOMAS J. LINSMEIER, DANIEL B. THORNTON, MOHAN VENKATACHALAM, and MICHAEL WELKER

The Effect of Mandated Market Risk Disclosures on Trading Volume Sensitivity to Interest Rate, Exchange Rate, and Commodity Price Movements 343

VENKY NAGAR

Delegation and Incentive Compensation 379

STEPHEN R. MOEHRLE

Do Firms Use Restructuring Charge Reversals to Meet Earnings Targets? 397

AUGUSTINE DURU and DAVID M. REEB

International Diversification and Analysts' Forecast Accuracy and Bias 415

APRIL KLEIN

Economic Determinants of Audit Committee Independence 435

WILLIAM H. BEAVER

Perspectives on Recent Capital Market Research 453

BOOK REVIEWS, Jane L. Reimers, Editor

Sunstein, Cass R., ed., *Behavioral Law and Economics* RICHARD B. DUSENBURY 475

Glover, Steven M., Stephen W. Liddle, and Douglas F. Prawitt, *E-Business: Principles and Strategies for Accountants* BRUCE C. BRANSON 476

Editorial Data 478

Editorial Policy and Style Information 479

Editorial Staff of *The Accounting Review*
2001–2002

EDITOR

Linda Smith Bamber, University of Georgia

CONSULTING EDITOR

E. Michael Bamber, University of Georgia

ASSOCIATE EDITORS

Greg Clinch, Melbourne Business School

Raffi J. Indjejikian, University of Michigan

Steven J. Kachelmeier, The University of Texas at Austin

Kenneth J. Klassen, University of Waterloo

Marlys Gascho Lipe, University of Oklahoma

Joan L. Luft, Michigan State University

Mark W. Nelson, Cornell University

Grace Pownall, Emory University

Richard C. Sansing, Dartmouth College

Daniel B. Thornton, Queen's University

Senyo Tse, The University of Texas at Austin

BOOK REVIEW EDITOR

Jane L. Reimers, Florida State University

American Accounting Association

EXECUTIVE DIRECTOR

Craig E. Polhemus

COMMUNICATIONS DEPARTMENT

Beverly Harrelson

Helen Griffith

Diane Hazard

Carol Stanton

Kathy Casper

Jan Kovarik (subcontracted)

OTHER AAA STAFF

Tracey E. Sutherland

Judy Cothorn

Debora Gardner

James Szelminski

Roland LaTulip

Linda Watterworth

Patricia A. Hoban

Diana Lombardi

Deirdre Harris

Barbara Brady

Mary Cole

Editorial Advisory and Review Board

Ashiq Ali, The University of Arizona

Shannon W. Anderson, Rice University

Benjamin C. Ayers, University of Georgia

Frances L. Ayres, University of Oklahoma

William R. Baber, The George Washington University

Stephen P. Baginski, Indiana University

Ramamurthy Balakrishnan, The University of Iowa

Orie E. Barron, Pennsylvania State University

Jan Barton, Emory University

Anne L. Beatty, Pennsylvania State University

Jacob G. Birnberg, University of Pittsburgh

Robert J. Bloomfield, Cornell University

Robert M. Bowen, University of Washington

Lawrence D. Brown, Georgia State University

David C. Burgstahler, University of Washington

Brian J. Bushee, University of Pennsylvania

Robert M. Bushman, The University of North Carolina at Chapel Hill

Carolyn M. Callahan, University of Arkansas

Youngsoo Susan Cheon, Chung-Ang University

Kevin C. W. Chen, Hong Kong University of Science & Technology

Chee W. Chow, San Diego State University

C. Bryan Cloyd, University of Illinois at Urbana-Champaign

Paul A. Copley, University of Georgia

Patricia M. Dechow, University of Michigan

Mark L. DeFond, University of Southern California

Ilia D. Dichev, University of Michigan

Sunil Dutta, University of California, Berkeley

Pieter T. Elgers, University of Massachusetts

Merle M. Erickson, University of Chicago

John (Harry) Evans III, University of Pittsburgh

Paul E. Fischer, Pennsylvania State University

Joseph G. Fisher, Indiana University

Jere R. Francis, University of Missouri–Columbia

O. Finley Graves, Kansas State University

Wayne R. Guay, University of Pennsylvania

David A. Guenther, University of Colorado at Boulder

Mahendra Gupta, Washington University

Susan F. Haka, Michigan State University

John R. Hand, The University of North Carolina at Chapel Hill

Donald R. Herrman, Oregon State University

D. Eric Hirst, The University of Texas at Austin

Chris E. Hogan, Southern Methodist University

Patrick E. Hopkins, Indiana University

K. E. (Skip) Hughes II, Louisiana State University

Christopher D. Ittner, University of Pennsylvania

Debra C. Jeter, Vanderbilt University

James J. Jambalvo, University of Washington

Deen Kemsley, Columbia University

S. Jane Kennedy, University of Washington

William R. Kinney, Jr., The University of Texas at Austin

Lisa L. Koonce, The University of Texas at Austin

William N. Lanen, University of Michigan

Charles M. C. Lee, Cornell University

Robert Libby, Cornell University

Robert C. Lipe, University of Oklahoma

Russell J. Lundholm, University of Michigan

Joseph Magliolo III, Southern Methodist University

Laureen A. Maines, Indiana University
Ella Mae Matsumura, University of Wisconsin–Madison
Steven R. Matsunaga, University of Oregon
Linda S. McDaniel, The University of North Carolina at Chapel Hill
Maureen F. McNichols, Stanford University
Richard R. Mendenhall, University of Notre Dame
Lillian F. Mills, The University of Arizona
James N. Myers, University of Illinois at Urbana–Champaign
Patricia C. O’Brien, University of Waterloo
Mark E. Peecher, University of Illinois at Urbana–Champaign
Kathy R. Petroni, Michigan State University
Jamie Pratt, Indiana University
Robert J. Ramsay, University of Kentucky
Madhav V. Rajan, University of Pennsylvania
Stephen G. Ryan, New York University
Katherine Schipper, Financial Accounting Standards Board

D. Shores, University of Washington
Paul J. Simko, Emory University
Richard G. Sloan, University of Michigan
J. Reed Smith, SUNY at Buffalo
Geoffrey B. Sprinkle, Indiana University
Thomas L. Stober, University of Notre Dame
Steve G. Sutton, University of Connecticut/University of Melbourne
Hun-Tong Tan, Nanyang Technological University
Siew Hong Teoh, The Ohio State University
Lynda Thoman, Purdue University
Richard Tubbs, The University of Iowa
Igor Vaysman, INSEAD
Robert E. Verrecchia, University of Pennsylvania
Martha L. Wartick, University of Northern Iowa
Gregory B. Waymire, Emory University
Hong Xie, University of Illinois at Urbana–Champaign
Teri L. Yohn, Georgetown University

PAST EDITORS

1926–1929	William A. Paton	1971–1972	Eldon S. Hendriksen
1929–1943	Eric L. Kohler	1973–1975	Thomas F. Keller
1944–1947	A. C. Littleton	1976–1978	Don T. DeCoster
1948–1949	Robert L. Dixon	1978–1983	Stephen A. Zeff
1950–1959	Frank P. Smith	1983–1987	Gary L. Sundem
1960–1962	Robert K. Mautz	1987–1990	William R. Kinney, Jr.
1963–1964	Lawrence L. Vance	1990–1994	A. Rashad Abdel-khalik
1965–1967	Wendell Trumbull	1994–1997	Robert P. Magee
1968–1970	Charles H. Griffin	1997–2000	Gerald L. Salamon

“Cost of Capital” in Residual Income for Performance Evaluation

Peter O. Christensen

University of Southern Denmark

Gerald A. Feltham

University of British Columbia

Martin G. H. Wu

University of Illinois at Urbana–Champaign

ABSTRACT: We consider a setting in which a firm uses residual income to motivate a manager's investment decision. Textbooks often recommend adjusting the residual income capital charge for market risk, but not for firm-specific risk. We demonstrate two basic flaws in this recommendation. First, the capital charge should not be adjusted for market risk. Charging a market risk premium results in “double” counting because a risk-averse manager will personally consider this risk. Second, while investors can avoid firm-specific risk through diversification, a manager cannot. If the manager faces significant firm-specific risk at the time he makes his investment decision, then it is optimal to charge him less than the riskless return so as to partially offset his reluctance to undertake risky investments. On the other hand, the manager will vary his investment decisions with the pre-decision information he receives, which accentuates his compensation risk, and the firm must compensate him for bearing this additional risk. Hence, if the manager will receive relatively precise pre-decision information, then it is optimal to charge him more than the riskless return to reduce the variability of his investment decisions.

We are grateful to participants at the UBCOW 2000 conference and research seminars at the University of California, Berkeley; University of Florida; University of Illinois at Urbana–Champaign; University of Minnesota; Queen's University; Stanford University; SUNY at Buffalo; the University of Technology-Sydney; and York University, as well as two anonymous reviewers and Raffi J. Indjejikian (associate editor), for helpful comments and suggestions. Detailed comments by Bjorn Jørgensen (discussant at Management Accounting Research Conference, 2001) and Stefan Reichelstein are gratefully acknowledged. Financial support has been provided by the Social Sciences Research Council of Denmark, the Association of State Authorized Public Accountants of Denmark, and the Social Sciences and Humanities Research Council of Canada.

*Submitted June 2000
Accepted May 2001*

Keywords: *cost of capital; residual income; management incentives; market risk.*

I. INTRODUCTION

Residual income is a performance measure that consists of some measure of operating income minus some charge for the capital used by the manager (or unit) being evaluated. The concept appeared as early as the 1920s (e.g., in DuPont's bonus plan calculation of its "Executive Trust Fund"), and has been frequently discussed in management accounting texts since General Electric adopted it in the 1950s. Despite its merits, residual income was not widely used in practice until recently. The business community's greater attention to this type of performance measure in recent years is due largely to the stimulus from Stern Stewart & Co.'s touting of their version of residual income, which they call Economic Value Added or EVA® (see Stewart 1991).¹

As a performance measure, residual income is designed to influence management's investment in capital assets, ideally inducing managers to undertake investments for which the net present value is positive and to reject those for which the net present value is negative. The rate used in calculating the capital charge, often called the "cost of capital," is clearly the riskless interest rate in a world of certainty or risk neutrality with no private pre-contract management information. However, we show that determining the appropriate measure of the cost of capital under uncertainty and risk aversion is complex, even if management does not have private pre-contract information.

In examining the impact of risk on the choice of the cost of capital used in incentive contracts, it is important to distinguish between the *firm-specific* and *market* components of the firm's risk. The firm-specific component is idiosyncratic and investors avoid it by diversifying. The investors share the market component, which covaries with the return on the market portfolio. From a market value perspective, firm-specific risk is irrelevant, whereas market risk is priced.²

Following the standard approach to valuation in the theory of finance, we adjust the expected outcome for risk, but most accounting professionals and empirical researchers risk-adjust the cost of capital. Therefore, it is not surprising that most management accounting textbooks state that the capital charge used in computing residual income should be the "cost of acquiring investment capital," such as the "weighted-average cost of capital" (WACC). For example:

Residual income = investment center's profit – (investment center's invested capital × imputed interest rate), where the imputed interest rate is the firm's cost of acquiring investment capital. (Hilton et al. 2000, 839)

To implement the residual income approach, corporate managers must specify an additional parameter, the risk-adjusted cost of capital for the division, which is then multiplied by division's net investment base to obtain a capital charge for the division. (Kaplan and Atkinson 1998, 506)

¹ Stern Stewart also propose the use of EVA® in valuing firms. Recent research in accounting (e.g., Ohlson 1995; Feltham and Ohlson 1995, 1996, 1999) has used residual income (sometimes referred to as "abnormal earnings") to characterize the relation between market value and accounting numbers.

² Feltham and Ohlson (1999) discuss the effect of market risk on the representation of market values, where market values are based on the current book value plus the discounted expected future residual income. As they demonstrate, the capital charge and discount rates should be based on the riskless interest rates, and the expected residual income numbers should be adjusted for market risk. This theoretically sound approach does not use a risk-adjusted cost of capital, which is commonly employed in practice and much of accounting research.

Corporate headquarters sets the rate used to calculate the capital charge. It should be higher than the corporation's rate for debt financing because the funds involved are a mixture of debt and higher cost equity. Usually, the rate is set somewhat below the company's estimated cost of capital . . . so the residual income of an average business unit will be above zero. (Anthony and Govindarajan 2000, 256)

Furthermore, WACC is an integral part of EVA[®] (see Stewart 1991, Chapter 12) and texts tend to reinforce this point when describing EVA[®] (e.g., Hilton et al. 2000, 841; Horngren et al. 2000, 830).

WACC may yield reasonably good decisions if the residual income numbers are forecasts that investors use to make centralized investment decisions. However, WACC is fundamentally flawed if residual income is the key performance measure used to evaluate a manager responsible for investment decisions. The flaw has two basic dimensions.

One dimension is that WACC ignores firm-specific risk even though that risk is likely to affect both a manager's investment decisions and the risk for which he must be compensated. If the manager's compensation varies with residual income, then, unlike investors, he cannot avoid his firm-specific risk (assuming he is precluded from going short in his own firm's stock). In our initial analysis of the effect of firm-specific risk on incentives we assume there is no market risk, so that WACC equals the riskless interest rate. We explore two extreme firm-specific information settings: no pre-decision information, and perfect pre-decision information. In both cases, the optimal cost of capital differs from the riskless interest rate, and is set by the owners so as to minimize the cost of underinvestment plus the cost of compensating the manager for firm-specific risk. The implications of this trade-off are quite different in the two extreme information settings. With no pre-decision information, the manager faces post-decision risk and, due to his risk aversion, charging him the riskless interest rate will induce him to underinvest (relative to the first-best level investors prefer). Reducing the manager's cost of capital induces more investment but increases the manager's compensation risk; the resulting trade-off makes it optimal to use a cost of capital strictly less than the riskless interest rate.³

On the other hand, with perfect pre-decision information there is risk with respect to the signal the manager may receive (e.g., it may be either good or bad news), but there is no post-decision risk. Consequently, charging the manager the riskless interest rate will induce him to vary his investment decision with the signal received and in each case he will choose the first-best level of investment. From an *ex ante* perspective (i.e., at the contract acceptance date), this increases the outcome risk relative to that in the no pre-decision information case. Increasing the manager's cost of capital will induce less investment for each signal and will reduce outcome variability, which in turn will reduce the manager's *ex ante* compensation risk; the resulting trade-off makes it optimal to use a cost of capital strictly greater than the riskless interest rate.⁴

The other dimension pertains to the issues associated with market risk. First, if a risk-averse manager cannot share the market risk in his compensation, then he will underinvest in projects with market risk (relative to the preferences of diversified investors) even if the

³ Dutta and Reichelstein (2002) obtain a similar result after they introduce agent risk aversion into their private pre-contract information setting. Giving the agent a "break" on the capital charge induces him to invest in a risky project. See also Lambert (2001).

⁴ Similar results are obtained in models with private pre-contract information and risk-neutral agents (see, e.g., Antle and Eppen 1985; Antle and Fellingham 1990; Dutta and Reichelstein 2002). There is no risk premium in these settings, but charging a cost of capital higher than the cost to the principal reduces the informed agent's expected *information rent*.

cost of capital equals the riskless interest rate. Using WACC accentuates this underinvestment, since WACC includes the investors' risk premium, and this effectively induces the manager to "double-count" the cost of market risk. Second, if the manager can trade in the shares of other firms, then he can use such trades to efficiently share the market risk inherent in his compensation. In that case, if there is no firm-specific risk and the cost of capital is set equal to the riskless interest rate (not WACC), then the manager will make precisely the same investment decisions as investors.

As briefly described above, our analysis establishes that management accounting professionals, researchers, and teachers should take considerable care in specifying the cost of capital that should be used in computing residual income for incentive purposes. Of course, our results depend, at least in part, on the specifics of our model.

We use a one-period agency model in which the manager receives private information after accepting a compensation contract, but before making his production choice. The production choice includes effort that is costly to the manager and an investment level that is costly to investors. The manager's effort is unobservable, so that the principal must impose incentive risk on the manager. This, plus the unobservability of the manager's pre-decision information, induces an incentive problem with respect to his signal-contingent investment choice even though the investment level is observable and not costly to the manager. The production model, which we describe in Section II, assumes that effort and investment have an additive effect on the outcome, and that uncertainty pertains to the productivity of the investment such that outcome risk increases with the level of investment. The random rate of investment productivity has firm-specific and market components.

To simplify the analysis and to create a clear role for residual income, we assume that the manager's incentive contract is restricted to be a linear function of residual income, which is equal to the outcome minus 1 plus the cost of capital times the level of investment. Section III introduces the manager's utility function and derives the production choices induced by a given incentive contract and each possible private signal. We then characterize the manager's *ex ante* contract acceptance decision.

We assume that the well-diversified investors who own the firm are represented by a principal who seeks to contract with the manager to maximize the firm's net market value. Section IV assumes there is no market risk and characterizes the linear residual income incentive contract that the principal will choose. The analysis initially assumes that the manager receives perfect productivity information before making his production choice and then assumes he receives no productivity information. In Section V we extend the no-pre-decision information case to consider the impact of market risk, assuming that the manager can trade in the market portfolio. Throughout Sections III, IV, and V, we assume that the capital provided by investors is the amount invested in the manager's risky project; e.g., the manager can neither invest in marketable securities nor supplement the investors' capital by borrowing. In Section VI we establish that relaxing that assumption implies that the optimal cost of capital equals the riskless interest rate even if there is firm-specific risk (since otherwise the manager has an arbitrage opportunity). In Section VII we provide some concluding remarks.

II. BASIC PRODUCTION MODEL

Our model has two factors of production: the effort supplied by the manager, denoted by its cost a , and the capital supplied by the principal, denoted by its cost q , with $a \in A = [0, \infty)$ and $q \in Q = [0, \infty)$. The outcome (e.g., net revenue) $\tilde{x}$ generated by these two factors also depends on a random productivity factor $\tilde{\delta}$. In particular, we assume:

$$\tilde{x} = h(a) + \tilde{\delta}g(q),$$

where $h(a)$ and $g(q)$ are increasing, concave production functions with $h(0) = g(0) = 0$ and infinite marginal productivities at $a = 0$ and $q = 0$. The additive separability between the two factors of production, a and q , simplifies the analysis, whereas the multiplicative relation $\tilde{\delta}g(q)$ is a simple means of having the investment choice q influence the riskiness of the outcome.⁵ We let $\tilde{\delta} = \bar{\delta} + \tilde{\epsilon}_\delta$, where $\bar{\delta}$ is the prior mean and $\tilde{\epsilon}_\delta \sim N(0, \tau_\delta^2)$ is the random component.

We consider the impact of market risk in Section V. However, in our basic model, we assume there is no market risk, so that the principal (representing well-diversified investors) is effectively risk neutral. The principal and manager sign a contract z at $t = 0$. The manager then receives private pre-decision information y about capital productivity, and makes his production decision $d = (a, q)$. The outcome $\tilde{x}$ is realized at $t = 1$.

Assuming the random variables are jointly normally distributed, we can let $\tilde{\epsilon}_\delta = \tilde{\epsilon}_y + \tilde{\epsilon}_p$, where $\tilde{\epsilon}_y \sim N(0, \tau_y^2)$ is the random component of productivity that is revealed by the manager's signal and $\tilde{\epsilon}_p \sim N(0, \tau_p^2)$ is the component that is not revealed until the outcome is realized, with $\text{Cov}[\tilde{\epsilon}_y, \tilde{\epsilon}_p] = 0$ and $\tau_\delta^2 = \tau_y^2 + \tau_p^2$. We can then let:

$$\tilde{y} \equiv \tilde{\delta} + \tilde{\epsilon}_y$$

and

$$\tilde{\delta} = y + \tilde{\epsilon}_p.$$

With this representation, the signal $y \in Y = \mathbb{R}$ equals the manager's posterior mean with respect to capital productivity, i.e., $E[\tilde{\delta}|y] = y$. We refer to τ_y^2 as a measure of the pre-decision information risk and τ_p^2 as a measure of the post-decision risk (with respect to capital productivity).

III. MANAGER'S DECISIONS

We assume the manager has the following exponential utility function with respect to his wealth w at date $t = 1$ and his effort a at date $t = 0$:

$$u(w, a) = -\exp[-b_a(w - Ra)], \quad (1)$$

where b_a is the manager's risk aversion and Ra is his personal effort cost expressed in units of terminal wealth. We assume the manager's time preference is the same as that implicit in the market's riskless return R , so that the investors and the manager are indifferent between paying (receiving) a dollar at $t = 0$ and R dollars at $t = 1$.⁶

In our basic model, with no market risk and manager utility function (1), the manager has no incentive to invest in the market, so that his terminal wealth w is equal to his

⁵ Many papers in the literature assume that the random component of the outcome is additive. However, Bushman et al. (2000) and Feltham and Wu (2001) consider models with multiplicative random productivity; they apply it to effort and do not consider capital investment.

In our analysis the manager can influence the firm's risk only by varying the capital investment, whereas Demski and Dye (1999) consider a setting in which the risk-return combination is a separate choice variable.

A reviewer suggested that we assume $\tilde{x} = h(a) + \tilde{\delta}g(q) + \tilde{\epsilon}_0$. This has the virtue of including a base level of uncertainty that does not disappear with $q = 0$ and, furthermore, implies that q and x do not reveal the manager's signal if the latter provides perfect pre-decision information about $\tilde{\delta}$. We have established that this would not materially affect our basic results and, hence, for simplicity we do not include $\tilde{\epsilon}_0$.

⁶ We could, without loss of generality, measure the units of current consumption so that $R = 1$, but that would tend to obscure the relation between the cost of capital and the riskless interest rate.

compensation c , which is paid at $t = 1$.⁷ The compensation depends on the contractible information at $t = 1$, which is assumed to consist of the outcome x and the investment q ; i.e., the principal does not observe the manager's action a and pre-decision information y . Furthermore, we initially assume the manager's investment in the firm's risky project must equal the capital q provided by the principal (i.e., he can neither invest some in the capital market, nor can he supplement q by borrowing).

In our single-period model, we define residual income to be $x - \Gamma q$, where Γ is the charge for using a dollar of the principal's capital, i.e., $\Gamma - 1$ is the endogenously determined "cost of capital." We restrict our analysis to linear compensation functions of residual income, which we express as:⁸

$$c = \alpha + \nu[x - \Gamma q].$$

Hence, the compensation contract is expressed as $z = (\alpha, \nu, \Gamma)$. We interpret α as the manager's fixed wage and ν as his incentive rate with respect to residual income.

Production Choice

In deriving the relation between the manager's production decision $d = (a, q)$ and contract z , we note that outcome $\tilde{x}$ is normally distributed (given signal y), which implies the manager's wealth (compensation), $\tilde{w}$, is normally distributed. Hence, we can express the manager's expected utility as a function of his certain-equivalent $CE(y, d, z)$, which is a function of his expected wealth and the variance of that wealth, minus the cost of effort:

$$E[u(\tilde{w}, a)|y, d, z] = -\exp[-b_a CE(y, d, z)], \quad (2a)$$

$$\text{where} \quad CE(y, d, z) = E[\tilde{w}|y, d, z] - \frac{1}{2}b_a \text{Var}[\tilde{w}|y, d, z] - Ra, \quad (2b)$$

$$E[\tilde{w}|y, d, z] = \alpha + \nu[h(a) + yg(q) - \Gamma q], \quad (2c)$$

$$\text{Var}[\tilde{w}|y, d, z] = \nu^2 g(q)^2 \tau_p^2. \quad (2d)$$

We assume the parameters are such that the manager's optimal production choice $d^+ = (a^+, q^+)$ given y and z is characterized by the following certainty-equivalent first-order conditions (we denote the optimal-choice variables with a dagger):

$$CE_a(y, d^+, z) = 0 \Leftrightarrow \nu h'(a^+) = R, \quad (3a)$$

$$CE_q(y, d^+, z) = 0 \Leftrightarrow yg'(q^+) = \Gamma + b_a \nu g'(q^+)g(q^+)\tau_p^2, \quad \text{if } y > 0, \quad (3b)$$

and $q^+ = 0$ if $y \leq 0$.

The additive separability of the production function with respect to a and q , and the fact that a does not influence the riskiness of output, imply the following (which are derived directly from Equation [3a]):

⁷ We assume the compensation contract precludes the manager from investing in the shares of his own firm in order to avoid informed insider trading and trading to undo incentive risk.

⁸ Observe that this contract is equivalent to a linear contract of the form $\alpha + \beta_x x + \beta_q q$, with $\nu = \beta_x$ and $\Gamma = -\beta_q/\beta_x$. Although a linear contract is not optimal, a restriction to linear contracts is common in the incentive literature (e.g., Bushman and Indjejikian 1993; Feltham and Xie 1994; Indjejikian and Nanda 1999; Bushman et al. 2000; Feltham and Wu 2000; Datar et al. 2001; Dutta and Reichelstein 2002). As in those papers, restricting the analysis to linear contracts significantly simplifies the analysis given that the performance measures are normally distributed and the manager has exponential utility. It has the additional benefit in this paper of permitting us to interpret the contract as a function of residual income.

Remark 1:

- (a) The manager's induced effort choice increases with the incentive rate on residual income, and is independent of the manager's pre-decision information and the capital charge used to compute residual income, i.e., $da^*/dv > 0$, $da^*/dy = 0$, and $da^*/d\Gamma = 0$.
- (b) For any action a there is a unique incentive rate v that will induce that action:

$$v^*(a) = R/h'(a). \quad (3a')$$

The manager's capital investment decision is determined by relation (3b), and the following remark characterizes some key implications of that relation:

Remark 2:

- (a) For $y > 0$, the manager's induced investment choice increases with the signal-contingent expected productivity of capital, and decreases with the incentive rate, the manager's risk aversion, his post-decision risk, and the capital charge, i.e., $dq^*/dy > 0$, $dq^*/dv < 0$, $dq^*/db_a < 0$, $dq^*/d\tau_p^2 < 0$, and $dq^*/d\Gamma < 0$.
- (b) If the manager does not receive any pre-decision information (i.e., $\tau_y^2 = 0$, $\tau_p^2 = \tau_\delta^2$, and $y = \delta$), then for any desired effort and investment levels a and q , there exists a unique capital charge:⁹

$$\Gamma^*(a, q) = [\bar{\delta} - b_a(R/h'(a))g(q)\tau_\delta^2]g'(q) \quad (3b')$$

that will induce the manager to select q .

To recognize explicitly that the investment choice characterized by relation (3b) depends on y , we denote that choice by $q^*(y)$. Similarly, $d^*: Y \rightarrow A \times Q$ specifies the manager's production decision for each possible signal (given contract z), where $d^*(y) \equiv (a^*, q^*(y))$, which recognizes that the choice of a is independent of y .

Contract Acceptance

The manager must decide whether to accept the contract the principal offers him before he receives his private pre-decision information (and he cannot renege after observing his signal). If the manager does not accept the contract, then he receives fixed compensation c^0 from alternative employment that does not require any effort, so that his reservation utility is:

$$U^0 \equiv -\exp[-b_a c^0]. \quad (4)$$

If the manager accepts contract z , then his expected utility at $t = 0$ is:

$$U^*(z) = E[-\exp[-b_a CE(\tilde{y}, d^*(\tilde{y}), z)]] = -\exp[-b_a CE^*(z)], \quad (5a)$$

$$\text{where} \quad CE^*(z) \equiv \alpha + v h(a^*) - R a^* - \frac{1}{b_a} \ln \{E[\exp[-b_a [f(\tilde{y}, v, \Gamma) - \pi^p(\tilde{y}, v, \Gamma)]]]\} \quad (5b)$$

is the manager's certainty equivalent at the time of contract acceptance,

⁹ We obtain this expression by substituting Equation (3a') for v in relation (3b) and solving for Γ .

$$f(y, \nu, \Gamma) \equiv \nu[yg(q^+(y)) - \Gamma q^+(y)] \quad (5c)$$

is the component of the manager's expected compensation that varies with y , and

$$\pi^p(y, \nu, \Gamma) \equiv \frac{1}{2} b_a \nu^2 g(q^+(y))^2 \tau_p^2$$

is the manager's post-decision risk premium given y . (Note: The above imply $f(y, \nu, \Gamma) = \pi^p(y, \nu, \Gamma) = 0$ if $y \leq 0$.)

Since the fixed wage α does not influence the manager's production decision, it is optimal for the principal to set it so that the manager is indifferent between accepting and not accepting the contract, i.e., $CE^+(z) = c^0$. This implies that the expected compensation cost to the principal is:

$$E^+[\tilde{c}|\nu, \Gamma] = c^0 + Ra^+ + \pi(\nu, \Gamma), \quad (6a)$$

$$\text{where} \quad \pi(\nu, \Gamma) \equiv E[f(\tilde{y}, \nu, \Gamma)] + \frac{1}{b_a} \ln \{E[\exp[-b_a[f(\tilde{y}, \nu, \Gamma) - \pi^p(\tilde{y}, \nu, \Gamma)]]]\} \quad (6b)$$

is the risk premium paid to the manager.

Observe that π represents the total premium for *ex ante* compensation risk, which is due to the variability in pre-decision productivity information (and the associated variability in investment) and the post-decision variability in productivity. In the next section, we consider the extreme settings in which there is either no post-decision risk or no pre-decision information.

IV. THE PRINCIPAL'S DECISION PROBLEM

We assume a well-diversified group of investors owns the firm. The firm's shares are publicly traded at the current date ($t = 0$) and the principal (e.g., the firm's board of directors) seeks to maximize the firm's current market value. In our basic model, with no market risk, the market value given contract z is:

$$V(d, z) = E[\beta(\bar{x} - \tilde{c}) - \tilde{q}|d, z],$$

where $\beta \equiv R^{-1}$ is the riskless discount factor. Our analysis assumes that the manager's compensation is a linear function of net income and a constant charge for using capital, i.e., $c = \alpha + \nu[x - \Gamma q]$.¹⁰

Setting α so that the manager's certainty equivalent equals c^0 and using the decision rule d^+ induced by ν and Γ , consistent with conditions (3a) and (3b), yield the following unconstrained optimization problem for the principal's choice of ν and Γ :

¹⁰ The presence of post-contract, pre-decision private information raises the question of whether it is optimal for the principal to offer the manager a contract that consists of a menu from which the manager could later choose by submitting a message m after he observes his private pre-decision information. For example, a menu of linear contracts, $c(x, q; m) = \alpha(m) + \nu(m)[x - \Gamma(m)q]$, where the contract parameters including the capital charge depend on the manager's message, may be valuable to the principal. Of course, there might exist an equivalent contract with no *direct* communication in which the contract parameters depend on the capital investment, i.e., $c(x, q) = \alpha(q) + \nu(q)[x - \Gamma(q)q]$. The key is whether the observable capital investment *indirectly* reveals the relevant aspects of the manager's private information (see, e.g., Melumad and Reichelstein 1987; Christensen and Feltham 1993, for settings in which this occurs). In any event, we restrict our analysis to linear contracts with constant contract parameters. This entails a loss of generality, but it provides a well-defined notion of a capital charge, i.e., a constant marginal cost to the manager for his use of the principal's capital.

$$\begin{aligned} \max_{\nu, \Gamma} V^*(\nu, \Gamma) = & \beta \{h(a^*) + E[\tilde{y}g(q^*(\tilde{y}))]\} - [c^0 + Ra^* + \pi(\nu, \Gamma)] \\ & - E[q^*(\tilde{y})]. \end{aligned} \quad (7)$$

That is, the optimal contract choice (ν^*, Γ^*) maximizes the current market value of the investors' gross pay-off $\tilde{x} - \tilde{q}$ minus the expected compensation (which is the sum of the manager's reservation wage, his effort costs, and his risk premium). Differentiating with respect to ν and Γ yields the first-order conditions that characterize the optimal contract choice (ν^*, Γ^*) . Details appear in the Appendix, along with the proofs of the following propositions.

To simplify the analysis and illustrate the importance of distinguishing between pre-decision information risk and post-decision risk, we now focus on two extreme information settings. In the first, the manager receives perfect pre-decision information (i.e., $y = \tilde{\delta} + \epsilon_{\delta}$, $\tau_y^2 = \tau_{\delta}^2$ and $\tau_p^2 = 0$), whereas in the second he does not receive any pre-decision information (i.e., $y = \tilde{\delta}$, $\tau_y^2 = 0$, and $\tau_p^2 = \tau_{\delta}^2$).

Perfect Pre-Decision Information

We now consider the extreme information setting in which the manager receives perfect pre-decision information, i.e., $\tilde{y} = \tilde{\delta}$. Hence, there is no post-decision risk, but there is information risk. Given contract $z = (\alpha, \nu, \Gamma)$, the lack of post-decision risk (i.e., $\tau_p^2 = 0$ in Equations [3a] and [3b]) implies the manager will choose a^* and $q^*(y)$ such that:

$$\nu h'(a^*) = R, \quad \text{all } y, \quad (8a)$$

$$\begin{cases} q^*(y) = 0, & \text{if } y \leq 0, \\ yg'(q^*(y)) = \Gamma, & \text{if } y > 0. \end{cases} \quad (8b)$$

Since there is no post-decision risk, the principal could induce the first-best production decision by setting $\nu = 1$ and $\Gamma = R$. However, due to the variability in the pre-decision information he might receive, the manager faces compensation risk at the time he accepts the contract, and he must be paid a premium for that risk. Observe that Equation (5b) implies $\pi^p = 0$ if $\tau_p^2 = 0$. Hence, Equation (6b) becomes:

$$\pi(\nu, \Gamma) = E[f(\tilde{y}, \nu, \Gamma)] + \frac{1}{b_a} \ln \{E[\exp[-b_a f(\tilde{y}, \nu, \Gamma)]]\}. \quad (9)$$

As demonstrated in the Appendix, if the manager is risk averse and there is firm-specific risk (which is resolved by y), this risk premium is increasing in ν and decreasing in Γ . This leads to the following result (see the Appendix for proofs).

Proposition 1: With perfect pre-decision information (i.e., $\tau_y^2 = \tau_{\delta}^2 > 0$ and $\tau_p^2 = 0$) and a risk-averse manager (i.e., $b_a > 0$), it is optimal for the principal to set the capital charge Γ^* *strictly greater* than the riskless return R , and the incentive rate ν^* *strictly less* than 1. If the manager is risk neutral (i.e., $b_a = 0$), then the first-best result is achieved with $\Gamma^* = R$ and $\nu^* = 1$.

There is no post-decision risk, but the outcome and compensation vary with the signal y , inducing *ex ante* compensation risk. Reducing ν does not affect the manager's investment

decision, but it makes his compensation less variable. On the other hand, increasing Γ induces the manager to invest less for any given $y > 0$. Reduced signal-contingent investments reduce the variability in residual income, which in turn reduces the variability of the manager's compensation, so that the risk premium paid to the manager is decreasing in Γ .

Models with pre-contract, private information and risk-neutral managers (see, e.g., Antle and Eppen 1985; Antle and Fellingham 1990; Dutta and Reichelstein 2002) obtain similar results (but for a very different reason). There is no risk premium in these settings, but the principal optimally induces underinvestment to curtail the informed manager's informational rents. In both types of model, a capital charge higher than the principal's cost of capital induces underinvestment relative to first-best.

No Pre-decision Information

We now consider the setting in which the manager does not receive any pre-decision information about δ . In that case, there is a unique optimal investment level q (as well as a unique optimal effort level a). From Remarks 1 and 2 we know that for any effort level a there is a unique incentive rate ν that will induce a , and for any investment level q (given the incentive rate ν) there is unique capital charge Γ that will induce q . Since q is contractible information, the desired investment level could be induced by specifying the amount that will be provided by the principal (essentially a penalty contract). However, it is instructive to identify the capital charge that will induce the desired investment choice if it is delegated to the manager, since we view this setting as the limiting case in which little of the manager's uncertainty is resolved before he makes his production decision. That is, post-decision risk is a central feature of this setting, and we are interested in identifying the impact of firm-specific post-decision risk on the optimal capital charge in the manager's incentive contract.

Given contract $z = (\alpha, \nu, \Gamma)$, the lack of pre-decision information (i.e., set $y = \bar{\delta}$ in Equations [3a] and [3b]) implies the manager will choose $a^\dagger$ and $q^\dagger$ such that:

$$\nu h'(a^\dagger) = R, \quad (10a)$$

$$\bar{\delta} g'(q^\dagger) = \Gamma + b_a \nu g'(q^\dagger) g(q^\dagger) \tau_p^2. \quad (10b)$$

Furthermore, the manager's risk premium is entirely attributable to post-decision risk, i.e.,

$$\pi(\nu, \Gamma) = \pi^p(\bar{\delta}, \nu, \Gamma) = \frac{1}{2} b_a \nu^2 g(q^\dagger)^2 \tau_p^2. \quad (11)$$

Condition (10a) implies that the principal can induce the first-best effort by setting the incentive rate equal to 1, and that decreasing the incentive rate ν reduces effort (i.e., $a_\nu^\dagger > 0$). The incentive rate ν has a direct and an indirect effect on the risk premium. The direct effect of decreasing ν is to reduce the risk premium due to reduced variability in the compensation for a given investment level. The indirect effect is to increase the risk premium due to increased investment, since Equation (10b) implies that reducing the incentive rate increases the capital investment (i.e., $q_\nu^\dagger < 0$). As demonstrated in the Appendix, the net result of these effects is that it is again optimal for the principal to set the incentive rate less than 1. The increase in investment and the direct reduction in the risk premium offset the resulting loss of efficient effort.

Unlike the perfect pre-decision information case, Equation (10b) implies that, due to post-decision risk, a capital charge Γ equal to the riskless return R will induce the manager to invest less than the first-best level if the incentive rate is positive. Furthermore, Equations

(10b) and (11) imply that increasing the capital charge decreases the capital investment (i.e., $q_r^+ < 0$) and thereby indirectly decreases the risk premium. As demonstrated in the Appendix, it is optimal for the principal to set the capital charge less than the riskless return. The expected increase in investment productivity more than offsets the increase in risk premium.¹¹

We summarize the above comments in the following proposition.

Proposition 2: With no pre-decision information (i.e., $\tau_p^2 = \tau_s^2 > 0$ and $\tau_y^2 = 0$) and a risk-averse manager (i.e., $b_a > 0$), it is optimal for the principal to set the capital charge Γ^+ *strictly less* than the riskless return R , and the incentive rate v^+ *strictly less* than 1. If the manager is risk neutral (i.e., $b_a = 0$), then the first-best result is achieved with $\Gamma^+ = R$ and $v^+ = 1$.

Propositions 1 and 2 establish that, with a risk-averse manager, both pre-decision information and post-decision risk make it optimal to set the incentive rate less than 1. However, pre-decision information makes it optimal to set the capital charge greater than the riskless return, whereas post-decision risk makes it optimal to set the capital charge less than the riskless return. In our general formulation, there are both pre-decision information and post-decision risk, and we provide general expressions for the optimal incentive rate and capital charge in the Appendix. Unfortunately, these are complex and it is difficult to make precise statements about the optimal incentive rate and capital charge. However, based on our results from the two extreme cases, it appears likely that the optimal incentive rate is always less than 1, whereas the sign of the difference between the optimal capital charge and the riskless return depends on whether the pre-decision information effect or the post-decision risk effect is dominant.

V. MARKET RISK

As discussed in the introduction, the management accounting literature has generally held that the appropriate "cost of capital" in measuring residual income is the "weighted average cost of capital." This implies that firm-specific (i.e., diversifiable) risk is irrelevant and that the focus is on the relation of the firm's outcome to the market return. In Section IV, we demonstrated that, in the absence of market risk, firm-specific risk affects the optimal capital charge if a principal uses residual income as a performance measure in a setting in

¹¹ In our linear contract analysis, we assume, for tractability, that there is no lower bound on compensation. However, with general compensation contracts there is no optimal contract due to the "Mirrlees (1975) problem" unless we restrict the class of feasible contracts to be uniformly bounded from below by some constant compensation $\underline{c}$. Within this class of contracts (and focusing on the setting with no pre-decision information), the optimal compensation is $\underline{c}$ for residual income $x - \Gamma q$ below some cutoff, and is an increasing *strictly concave* function of residual income above the cut-off. If $\underline{c}$ is sufficiently small (so that it has only a small probability of occurrence), then it is our conjecture that the qualitative characteristics of the linear contract results will hold with an optimal bounded contract in which the principal uses a capital charge to induce the manager to select the optimal capital investment. That is, if the capital charge is equal to the riskless return, then the manager will invest less than the principal prefers (even less than with linear contracts due to the "concavity" of the contract) and, hence, it will be optimal for the principal to set the capital charge less than the riskless return.

It is quite possible the result would change if $\underline{c}$ is sufficiently high that the "option" aspect of the compensation plays a major role. Risk is independent of effort, but it increases with an increase in the capital investment. The optimal bounded contract for inducing effort and a capital charge equal to the riskless return may induce the manager to invest more than the principal prefers (since the expected utility of an option increases with risk if the option is not "too far in-the-money") and, hence, it would be optimal for the principal to set the capital charge greater than the riskless return. However, the lack of closed form solutions for the optimal bounded contract makes it difficult to make precise statements about the magnitude of these option effects.

which capital investment affects the amount of firm-specific risk. Interestingly, in that setting, firm-specific risk does not affect the first-best investment level, but it does affect the second-best level.

We now introduce market risk into the no-pre-decision-information setting. As we demonstrate, market risk influences both the first- and second-best investment levels, *but it does not directly affect the optimal capital charge in the manager's performance measure*. This result depends crucially on the assumption that either the manager can invest (go long or short) in the market portfolio or the principal can vary the manager's compensation with the return on the market portfolio. The effects of firm-specific risk are essentially the same as in Section IV, in which there was no market risk.

To introduce market risk, we assume the market portfolio yields a return of:

$$\tilde{R}_m = \bar{R}_m + \tilde{\epsilon}_m,$$

where $\tilde{\epsilon}_m \sim N(0, \sigma_m^2)$ is the random variation in the market return. The firm is owned by well-diversified investors who publicly trade the firm's shares at the current date ($t = 0$). The principal seeks to maximize the firm's current market value, which is assumed to take the following form:¹²

$$V(d, z) = \beta \{E[\tilde{x} - \tilde{c}|d, z] - b_m \text{Cov}[\tilde{x} - \tilde{c}, \tilde{\epsilon}_m|d, z]\} - q,$$

where b_m can be interpreted as a measure of the investors' aggregate risk aversion.

The variation in the firm's investment productivity is divided into firm-specific and market components, represented by $\tilde{\epsilon}_s = \tilde{\epsilon}_s + \zeta \tilde{\epsilon}_m$, where $\tilde{\epsilon}_s \sim N(0, \tau_s^2)$ is the firm-specific variation and ζ is a covariance parameter (which can be positive or negative). These random variables are such that $\text{Cov}[\tilde{\epsilon}_s, \tilde{\epsilon}_m] = 0$ and $\text{Cov}[\tilde{\epsilon}_s, \tilde{\epsilon}_m] = \zeta \sigma_m^2$.

¹² This is equilibrium market price in an economy in which the terminal values of all firms are normally distributed, investors have negative exponential utility, and b_m is the normalized measure of the investors' aggregate risk aversion (see, e.g., Rubinstein 1973; Ohlson 1990). More specifically, if V_j and X_j represent the current and terminal values of firm j , and b_i is the risk aversion of investor i (so that b_i^{-1} is the corresponding risk tolerance), then $b_m = \sum_i b_i / V_m$, where $V_m = \sum_j V_j$. Furthermore, $x_m = \sum_j X_j / V_m \sim N(\bar{R}_m, \sigma_m^2)$.

There is no concept of a "risk-adjusted cost of capital" in this value relation. The riskless discount factor β is applied to the risk-adjusted expected terminal value of the firm. This is a consumption-based asset pricing model, but in our simple model aggregate consumption equals the terminal value of all firms.

Although there is no fundamental role for a "risk-adjusted cost of capital" in this analysis, we can restate the firm's current market value as:

$$V(d, z) = \hat{\beta}(d, z) E[\tilde{x} - \tilde{c}|d, z],$$

where

$$\hat{\beta}(d, z) = \hat{R}(d, z)^{-1}$$

and

$$\hat{R}(d, z) = R + b_m \text{Cov}[\tilde{x} - \tilde{c}, \tilde{\epsilon}_m|d, z] / V(d, z).$$

That is, we can interpret $\hat{R}(d, z) - 1$ as the firm's required expected return in equilibrium given the market risk of its terminal value, i.e., the "risk-adjusted cost of capital." Hence, if we consider a setting in which the firm uses a hurdle rate for investment project approval and there is no incentive problem, then the rate of return of investments must be above the risk-adjusted cost of capital to have positive net present value. However, we need to calculate $V(d, z)$ to determine the risk-adjusted cost of capital, and, as indicated by the notation, $\hat{R}(d, z)$ potentially depends on the investments in the factors of production.

In the first-best setting, in which the manager receives a fixed wage $c = c^o + Ra$, the firm's current market value is:

$$V^*(a, q) = \beta \{ [h(a) + \delta g(q)] - [c^o + Ra] - b_m \zeta \sigma_m^2 g(q) \} - q.$$

Hence, the first-best production decisions are characterized by the following first-order conditions:

$$h'(a^*) = R,$$

$$\hat{\delta} g'(q^*) = R,$$

where

$$\hat{\delta} = \bar{\delta} - b_m \zeta \sigma_m^2$$

is the *risk-adjusted* expected productivity parameter for capital.¹³ Thus, for the investors, the risk adjustment of the productivity of capital is negative (positive) if the correlation between the firm's outcome and the return on the market portfolio is positive (negative), and it is not influenced by firm-specific risk.

In the second-best setting, the contractible information now includes the market return $\bar{R}_m$, so we expand the compensation relation to include this information at the rate λ_c :

$$\tilde{c} = \alpha + v[\bar{x} - \Gamma q] + \lambda_c \bar{R}_m.$$

The manager can personally invest (or go short) in the market portfolio, but not in any other portfolio. This has two quite distinct implications. First, he cannot go short in the shares of his own firm and, hence, he is precluded from "undoing" his incentive risk. Second, trading in the market portfolio allows him to partially offset "unnecessary" risk in a way that does not undermine his incentives, but reduces the risk premium he must be paid.¹⁴ Let λ_a represent the amount the manager personally borrows (at the riskless interest rate $R - 1$) for investment in the market portfolio, so that his terminal wealth is:

$$\tilde{w} = \alpha + v[\bar{x} - \Gamma q] + \lambda_c \bar{R}_m + \lambda_a [\bar{R}_m - R].$$

Let $d \equiv (a, q, \lambda_a)$ and $z \equiv (\alpha, v, \Gamma, \lambda_c)$ represent the elements of the manager's decision and compensation contract, respectively. Introducing market risk into Equation (2), with no pre-decision information, yields:

$$E[u(\tilde{w}, a) | d, z] = -\exp[-b_a CE(d, z)], \quad (12a)$$

¹³ If the manager receives firm-specific signal y before making his production decision, then $\hat{y} = y - b_m \zeta \sigma_m^2$ replaces $\hat{\delta}$.

¹⁴ Our approach is similar to that found in papers such as Leland and Pyle (1977) and Hughes (1986, 1988), and permits us to focus on distinguishing "priced and insurable" market risk vs. "unpriced and uninsurable" firm-specific risk. The ability to trade in the market portfolio partially removes unnecessary risk, but not as fully as permitting a broader range of trading. See Atkinson and Feltham (1981) for an analysis that considers these types of investments.

$$\text{where } CE(d,z) = E[\tilde{w}|d,z] - \frac{1}{2}b_a \text{Var}[\tilde{w}|d,z] - R_a, \quad (12b)$$

$$E[\tilde{w}|d,z] = \alpha + \nu[h(a) + \bar{\delta}g(q) - \Gamma q] + [\lambda_a + \lambda_c]\bar{R}_m - \lambda_a R, \quad (12c)$$

$$\text{Var}[\tilde{w}|d,z] = \nu^2 g(q)^2 [\tau_s^2 + \zeta^2 \sigma_m^2] + [\lambda_a + \lambda_c]^2 \sigma_m^2 + 2\nu g(q)[\lambda_a + \lambda_c]\zeta \sigma_m^2. \quad (12d)$$

Consequently, the first-order conditions characterizing the manager's choice of d are:

$$CE_a(d,z) = 0 \Leftrightarrow \nu h'(a) = R, \quad (13a)$$

$$CE_q(d,z) = 0 \Leftrightarrow \bar{\delta}g'(q) = \Gamma + b_a\{\nu g'(q)g(q)[\tau_s^2 + \zeta^2 \sigma_m^2] + g'(q)[\lambda_a + \lambda_c]\zeta \sigma_m^2\}, \quad (13b)$$

$$CE_\lambda(d,z) = 0 \Leftrightarrow \bar{R}_m - R = b_a[(\lambda_a + \lambda_c)\sigma_m^2 + \nu g(q)\zeta \sigma_m^2]. \quad (13c)$$

Let $\lambda \equiv \lambda_a + \lambda_c$ represent the manager's total "investment" in the market portfolio. From Equation (13c) it follows that this investment is a function of the incentive rate ν and his investment choice q :

$$\lambda^\dagger(\nu, q) = \lambda^\circ - \nu g(q)\zeta, \quad (13c')$$

where
$$\lambda^\circ \equiv \frac{b_m}{b_a}$$

is the investment the manager would make in the market portfolio if his compensation were independent of the market return and $\bar{R}_m - R = b_m \sigma_m^2$ is the risk premium per dollar invested in the market portfolio.¹⁵ The total investment λ is independent of λ_c , and we observe from Equation (12) that the only impact of λ_c is to reduce the cost to the manager of acquiring λ (i.e., to reduce $\lambda_a R$). However, the principal can also accomplish this by increasing the fixed wage α . Hence, if the manager can personally go long or short in the market portfolio, then there is no loss of generality if we assume the compensation function is such that $\lambda_c = 0$, which implies $\lambda = \lambda_a$.

The influence of the incentive rate ν and investment q on λ follows from the fact that the manager's compensation indirectly imposes market risk on him, and he uses λ to adjust his share of market risk so that it is efficient. As noted in the following remark, the nature of the "adjustment" depends on the nature of the covariance of the firm's outcome with the market.

Remark 3: The manager's investment in the market portfolio decreases with (is independent of, or increases with) the incentive rate if the covariance with the market is positive (zero, or negative), i.e., $d\lambda^\dagger/d\nu < (=, >) 0$ if $\zeta > (=, <) 0$.

Observe that the manager will "go short" in the market portfolio if $\nu g(q)\zeta > \lambda^\circ$, which is very possible.

¹⁵ This is consistent with the well-known result that if all investors have exponential utility, then they will hold the market portfolio in proportion to their risk tolerances, b_i^{-1} (see, e.g., Wilson 1968).

Substituting Equation (13c') into Equations (12c) and (12d) yields:

$$E[\tilde{w}|d,z] = \alpha + \nu[h(a) + \hat{\delta}g(q) - \Gamma q] + \lambda^0 b_m \sigma_m^2, \quad (12c')$$

$$\text{Var}[\tilde{w}|d,z] = \nu^2 g(q)^2 \tau_s^2 + \lambda^0 \sigma_m^2. \quad (12d')$$

Recall that $\hat{\delta}$ is the risk-adjusted expected productivity of capital. The manager can offset the market risk associated with the capital investment by investing in the market portfolio, but that is costly. Hence, the expected productivity of capital is adjusted for market risk. The firm-specific risk, on the other hand, cannot be avoided and, hence, the manager is left with a variance component, $\nu^2 g(q)^2 \tau_s^2$. We can now simplify the manager's decision problem for choosing a and q as:

$$\underset{a,q}{\text{maximize}} \alpha + \nu[h(a) + \hat{\delta}g(q) - \Gamma q] - \frac{1}{2}b_a \nu^2 g(q)^2 \tau_s^2 + \frac{1}{2}\lambda^0 b_m \sigma_m^2 - Ra,$$

and the first-order conditions are

$$\nu h'(a^+) = R, \quad (14a)$$

$$\hat{\delta}g'(q^+) = \Gamma + b_a \nu g'(q^+)g(q^+)\tau_s^2. \quad (14b)$$

Observe that Equation (14a) is identical to Equation (10a), whereas Equation (14b) differs from Equation (10b) due only to the market-risk adjustment of the expected value of δ (and the use of τ_s^2 instead of τ_p^2).¹⁶ Consequently, the introduction of market risk does not materially affect Remarks 1 and 2.

If the manager does not accept the contract offered to him by the principal, then he, again, not only receives fixed compensation c^0 from alternative employment, but also personally borrows λ^0 at the riskless interest rate and invests it in the market portfolio. Hence, his reservation utility and certainty equivalent are:¹⁷

$$U^0 = -\exp[-b_a CE^0], \quad (15a)$$

$$\text{where} \quad CE^0 \equiv c^0 + \frac{1}{2}\lambda^0 b_m \sigma_m^2. \quad (15b)$$

We follow an approach that is similar to that in Section IV to characterize the principal's choice of ν and Γ . We use the participation constraint (based on Equations [15a] and [15b]) to specify the fixed wage α , Equation (13c') to specify the manager's market portfolio choice of λ , and first-order conditions (14a) and (14b) to specify the effort a^+ and capital investment q^+ induced by incentive rate ν and capital charge Γ . The risk premium that must be paid to the manager is:

$$\pi(\nu, \Gamma) = \frac{1}{2}b_a \nu^2 g(q^+)^2 \tau_s^2, \quad (16)$$

which is the same as Equation (11) with $\tau_p^2 = \tau_s^2$. Of course, Equation (16) depends

¹⁶ If the manager receives firm-specific signal y , then τ_p^2 is retained, and $\hat{y} = y - b_m \zeta \sigma_m^2$ replaces $\hat{\delta}$.

¹⁷ The manager's wealth is $\tilde{w} = c^0 + \lambda_a [R_m - R]$ and, hence, his certainty equivalent is $CE(\lambda_a) = c^0 + \lambda_a b_m \sigma_m^2 - \frac{1}{2} b_a \lambda_a^2 \sigma_m^2$. Differentiating with respect to λ_a yields first-order condition $\lambda_a = \lambda^0$, so that $CE^0 = c^0 + \frac{1}{2}\lambda^0 b_m \sigma_m^2$.

crucially on the assumption that the manager can personally offset the market-risk component of his compensation by appropriately investing in the market portfolio. Consequently, we can express the principal's decision problem as:

$$\max_{v, \Gamma} V^+(v, \Gamma) = \beta \{h(a^+) + \delta g(q^+) - [c^0 + Ra^+ + \pi(v, \Gamma)]\} - q^+. \quad (17)$$

This is similar to Equation (7) for the no-pre-decision-information case. Hence, the characteristics of the optimal incentive rate v^+ and capital charge Γ^+ are essentially the same as in Proposition 2. More importantly, this establishes that market risk has no direct effect on the optimal capital charge.

Earlier in this section, we established that making the compensation contract contingent on the return on the market portfolio is irrelevant if the manager can personally invest in the market portfolio. We now consider the nature of the optimal contract $z = (\alpha, v, \Gamma)$ if the manager cannot invest in the market portfolio. Interestingly, if the principal can contract on $\tilde{R}_m$, then he will choose $\lambda_c^+ = \lambda^0 - v^+ g(q^+) \zeta$, which is precisely the same as the manager's choice (see Equation [14c']). Hence, it is irrelevant whether the manager can invest in the market portfolio or the principal can offer a relative-performance contract based on $\tilde{R}_m$.¹⁸

Proposition 3: Market risk has no direct effect on the optimal incentive rate or capital charge (in the no-pre-decision-information case) if the manager can trade in the market portfolio or the principal and manager can contract on the market return. Furthermore, $v = 1$ and $\Gamma = R$ if there is no firm-specific risk (i.e., $\tau_s^2 = 0$).

Observe that we do not claim that market risk has no effect on the incentive rate or capital charge. It is merely that the effect is indirect (and nonexistent if there is no firm-specific risk). This indirect effect arises because the principal's choice of q^+ to be induced is influenced by the market risk, and q^+ influences the firm-specific risk imposed on the manager.

If the manager cannot trade in the market portfolio and the principal and manager cannot contract on the market return, then the solution is more complex. Nonetheless, we again get the result that the optimal capital charge is less than the riskless return, except in a knife-edge case in which $v = 1$ (see the Appendix). In this setting, the contract is the only means for investors to "share" the market risk with the manager. Hence, we cannot rule out the possibility that the optimal incentive rate is greater than 1.¹⁹

Proposition 4: With market risk, no pre-decision information, and a risk-averse manager, it is optimal for the principal to set the capital charge Γ^+ *strictly less* than the riskless return R .

¹⁸ Fischer (1999) obtains a similar result in a setting in which there is no investment and no firm-specific risk, but "systematic" risk has a component that cannot be directly shared with the market as well as a component that can be shared (or what he calls "managed"). He also considers market trades as a form of earnings management by the firm. As discussed in Section VI, we preclude such trades because they would provide arbitrage opportunities for the manager given that the principal supplies capital in our setting.

¹⁹ From the proof in the Appendix, we observe that $v < (=, >) 1$ if $\frac{1}{2}b_a g(q^+)^2 \tau_q^2 > (=, <) \frac{1}{2}b_m \zeta \sigma_m^2 g(q^+)$, where the first term is the risk premium required by the manager and the second term is the risk premium required by investors.

The analysis in this section has examined the impact of market risk in the no-pre-decision-information setting. The effects of market risk in the perfect pre-decision information setting (see the subsection "Perfect Pre-Decision Information" in Section IV) are similar. That is, Proposition 1 continues to hold and market risk has no direct effect on the optimal capital charge.

VI. FIRM BORROWING AND LENDING

Up to this point, we have assumed that the manager must invest all of the capital input q from the principal in a single risky project. The manager can choose the amount invested but he has had no discretion as to where it is invested, nor can he supplement that investment by having the firm borrow. We now briefly consider a setting in which there is no pre-decision information, no market risk, and $q = q_f + q_r$, where the two components represent the amounts invested in the firm's risky project and the riskless asset, respectively.²⁰ The total capital q provided by the principal is contractible, but the allocation is not.²¹

In this setting, the firm's terminal dividend is:

$$\tilde{x} = h(a) + \tilde{g}(q_f) + Rq_r.$$

If the compensation contract is $c(x, q) = \alpha + \nu[x - \Gamma q]$ and the manager can personally invest in the market portfolio, then his certainty equivalent is:

$$CE(a, q) = E[\tilde{w}|a, q] - \frac{1}{2}b_a \text{Var}[\tilde{w}|a, q] - Ra, \quad (18a)$$

$$E[\tilde{w}|a, q] = \alpha + \nu[h(a) + \bar{g}(q_f) + Rq_r - \Gamma(q_f + q_r)], \quad (18b)$$

$$\text{Var}[\tilde{w}|a, q] = \nu^2 g(q_f)^2 \tau_p^2, \quad (18c)$$

where

$$q = (q_f, q_r).$$

From Equation (18b) it is obvious that the manager has an arbitrage opportunity if he can have the firm borrow or invest in the riskless asset and $\Gamma \neq R$. If $\Gamma < R$ (as in the no-pre-decision-information setting), then the manager will obtain as much capital q from the principal as possible and invest in the risky project as if $R - 1$ is his "cost of capital" (since it is his opportunity cost), i.e., as in Equation (10b) he will choose q_f such that:

$$\bar{g}'(q_f) = R + b_a \nu g'(q_f) g(q_f) \tau_p^2. \quad (19)$$

He will invest the remainder, $q - q_f$, in the riskless asset (at a profit to the manager).

We obtain similar results with $\Gamma > R$ (as in the perfect-pre-decision-information setting), but in that case the manager makes q as small as possible (even negative) and uses riskless borrowing as a means of financing q_f as specified in Equation (8b). However, there is a difference compared to the case in which $\Gamma < R$, in the sense that with $\Gamma > R$ the

²⁰ We could also consider a setting in which there is pre-decision information or market risk and the firm can go short or long in the market portfolio, but the basic result would be the same.

²¹ Of course, if the allocation is contractible and the principal can use a separate capital charge for the different types of investments, then the results of the prior analysis would not change. In this setting, the optimal capital charge for investments (or borrowing) in both the riskless asset and the market portfolio is the riskless interest rate, and the manager would have no incentive to have the firm make investments in those assets.

principal can prevent the manager from making arbitrage profits at the owners' expense by restricting the capital provided to be non-negative (i.e., initial dividends will not be rewarded at the rate Γ). In that case, the manager will not obtain capital from the principal, but will have the firm borrow at the riskless interest rate, which is a cheaper source of financing (or, equivalently, he will use leasing instead of acquiring the real assets used in production). In any case, the manager will invest in the risky project as if his cost of capital were the riskless interest rate.

From the above discussion it is obvious that if the principal cannot constrain how the manager uses the funds he supplies or how the manager finances risky investments, then the principal is effectively constrained to set the capital charge equal to the riskless return. This will lead to under- or overinvestment in the risky project relative to a setting in which a borrowing/lending constraint is feasible.

VII. CONCLUDING REMARKS

Conventional wisdom in the accounting literature argues that the "cost of capital" used to measure residual income should be adjusted for market risk. Risk-adjusting the cost of capital may be reasonable when one uses predicted residual income for valuation or centralized investment approval purposes. However, this approach is fundamentally flawed when the firm uses realized residual income as a performance measure to influence a manager's *decentralized* capital investment decisions.

There are two basic dimensions to the flaw. First, a manager is generally risk averse, and considers his personal risk when selecting his actions. Furthermore, a manager is generally more risk averse relative to the firm's risk than a typical well-diversified shareholder would be. Hence, he is likely to underinvest in risky projects, and charging a risk premium for the capital used will accentuate that underinvestment. Second, the manager can offset the market-risk component of his compensation by modifying his investment portfolio, or the principal can eliminate that risk by using a relative performance measure. As a result, market risk need not be a key factor in determining the cost of capital used in computing residual income for performance evaluation. Instead, the key factors are the impact of firm-specific risk on a manager's decisions and the risk premium he must be paid for the firm-specific risk he bears.

Our analysis focuses on two polar cases: perfect pre-decision information and no pre-decision information. In the first case, risk is not an issue when the manager selects his actions, and the principal can induce the first-best level of effort and capital investment by setting the incentive rate equal to 1 and the cost of capital equal to the riskless interest rate. However, when the manager accepts the contract, he anticipates the compensation risk that is attributable to both the variability in his anticipated pre-decision information and to the signal-contingent investment levels he will choose. The larger the investment levels induced, the greater the manager's compensation risk, and, hence, the greater the risk premium he must be paid. In this setting it is optimal to set the cost of capital greater than the riskless interest rate so as to induce an investment level less than first-best for each possible signal, and thereby to reduce the manager's firm-specific compensation risk. This, of course, assumes the principal can constrain the manager from riskless borrowing or going short in the market portfolio on the firm's account.

In the second case, firm-specific post-decision risk induces the manager to underinvest so as to partially avoid that risk. Setting the cost of capital less than the riskless interest rate will help mitigate the underinvestment problem. This will increase the agent's compensation risk, but the net effect is such that it is optimal to set the cost of capital strictly

less than the riskless interest rate (assuming the manager is constrained to invest all capital provided in the risky project).

These results indicate that in constructing performance measures, accountants (practitioners, researcher, and teachers) should carefully consider the impact of the manager's risk aversion and the differential impact of firm-specific risk vs. market risk. For example, in a recent empirical study of how firms tie CEO compensation to firms' stock market performance, Jin (2001, Abstract) found that "market risk was not necessarily a cost of providing incentives. . . . [T]he cost of giving incentives is the potential loss of diversification for CEOs. CEO incentive decreases with firm-specific risk, but may not decrease with market risk."

APPENDIX

General Contract Characterization

Differentiating Equation (7) with respect to ν and Γ yields:

$$\beta\{[h'(a^+) - R]a_v^+ + E[\tilde{y}g'(q^+(\tilde{y}))q_f^+(\tilde{y})] - \pi_v(\nu, \Gamma)\} - E[q_v^+(\tilde{y})] = 0, \quad (A1a)$$

$$\beta\{E[\tilde{y}g'(q^+(\tilde{y}))q_f^+(\tilde{y})] - \pi_\Gamma(\nu, \Gamma)\} - E[q_f^+(\tilde{y})] = 0, \quad (A1b)$$

with $q_v^+(y) = q_\Gamma^+(y) = 0$ for $y \leq 0$. Substituting Equation (3a) into Equation (A1a) and rearranging terms give:

$$\nu = 1 - \frac{\pi_v(\nu, \Gamma) - E[\{\tilde{y}g'(q^+(\tilde{y})) - R\}q_v^+(\tilde{y})]}{h'(a^+)a_v^+}. \quad (A2a)$$

Similarly, substituting Equation (3b) into Equation (A1b) gives

$$\Gamma = R + \frac{\pi_\Gamma(\nu, \Gamma) - E[b_a \nu g'(q^+(\tilde{y}))g(q^+(\tilde{y}))q_f^+(\tilde{y})\tau_p^2]}{E[q_f^+(\tilde{y})]}. \quad (A2b)$$

Proof of Proposition 1

With no decision risk, $\tau_p^2 = 0$ and Equation (8b) implies $q_v^+(y) = 0$, which in turn imply:

$$\nu = 1 - \frac{\pi_v(\nu, \Gamma)}{h'(a^+)a_v^+}, \quad (A2a')$$

$$\Gamma = R + \frac{\pi_\Gamma(\nu, \Gamma)}{E[q_f^+(\tilde{y})]}. \quad (A2b')$$

We have $h' > 0$ by assumption, and totally differentiating Equation (8a) establishes that $a_v^+ > 0$. Hence, Equation (A2a') implies $\nu < 1$, since

$$\begin{aligned} \pi_v(\nu, \Gamma) &= E[f_v(\tilde{y}, \nu, \Gamma)] - \frac{E[f_v(\tilde{y}, \nu, \Gamma) \exp[-b_a f(\tilde{y}, \nu, \Gamma)]]}{E[\exp[-b_a f(\tilde{y}, \nu, \Gamma)]]} \\ &= - \frac{\text{Cov}[f_v(\tilde{y}, \nu, \Gamma), \exp[-b_a f(\tilde{y}, \nu, \Gamma)]]}{E[\exp[-b_a f(\tilde{y}, \nu, \Gamma)]]} > 0. \end{aligned} \quad (A3)$$

The negative covariance follows from the fact that, for $y > 0$, $f_y = v[g + (yg' - \Gamma)q_y^+]$ $= v g > 0$, by Equation (8b), which implies $f_v = f/v$ is non-decreasing with y and $\exp[-b_a f]$ is non-increasing with y .

Totally differentiating Equation (8b) establishes that $q_R^+(y) < 0$ for $y > 0$, so that $E[q_R^+(\tilde{y})] < 0$. Hence, (A2b') implies $\Gamma > R$ since

$$\begin{aligned}\pi_r(v, \Gamma) &= E[f_r(\tilde{y}, v, \Gamma)] - \frac{E[f_r(\tilde{y}, v, \Gamma) \exp[-b_a f(\tilde{y}, v, \Gamma)]]}{E[\exp[-b_a f(\tilde{y}, v, \Gamma)]]} \\ &= - \frac{\text{Cov}[f_r(\tilde{y}, v, \Gamma), \exp[-b_a f(\tilde{y}, v, \Gamma)]]}{E[\exp[-b_a f(\tilde{y}, v, \Gamma)]]} < 0.\end{aligned}\quad (A4)$$

The positive covariance follows from the fact that both $\exp[-b_a f]$ and f_r are non-increasing with y . The former is proven above and the latter holds because, for $y > 0$, $f_r = v\{[yg - \Gamma]q_R^+ - q^+\} = -q^+$ (from Equation [8b]), which implies $f_{ry} = -q_y^+ < 0$.

Q.E.D.

Proof of Proposition 2

Differentiating Equation (11) with respect to v and Γ yields:

$$\pi_v(v, \Gamma) = b_a[v g(q^+)^2 + v^2 g'(q^+) g(q^+) q_v^+ \tau_p^2], \quad (A5a)$$

$$\pi_r(v, \Gamma) = b_a v^2 g'(q^+) g(q^+) q_r^+ \tau_p^2. \quad (A5b)$$

With no information risk, $y = \bar{\delta}$ from Equation (A2), we obtain:

$$v = 1 - \frac{\pi_v(v, \Gamma) - \{\bar{\delta} g'(q^+) - R\} q_v^+}{h'(a^+) a_v^+}, \quad (A2a'')$$

$$\Gamma = R + \frac{\pi_r(v, \Gamma) - b_a v g'(q^+) g(q^+) q_r^+ \tau_p^2}{q_r^+}. \quad (A2b'')$$

Substituting Equation (A5b) into Equation (A2b'') yields:

$$\Gamma = R - (1 - v) v b_a g'(q^+) g(q^+) \tau_p^2. \quad (A2b''')$$

Substituting Equations (A5a) and (10b) into Equation (A2a'') and then substituting Equation (A2b''') yield

$$\begin{aligned}v &= 1 - \frac{b_a[v g(q^+)^2 + v^2 g'(q^+) g(q^+) q_v^+ \tau_p^2] + [R - \Gamma - b_a v g'(q^+) g(q^+) \tau_p^2] q_v^+}{h'(a^+) a_v^+}, \\ &= 1 - \frac{b_a v g(q^+)^2 \tau_p^2}{h'(a^+) a_v^+} \leq 1.\end{aligned}\quad (A2a''')$$

If $b_a \tau_p^2 > 0$, then Equation (A2a''') implies $v < 1$ and, consequently, Equation (A2b''') implies $\Gamma < R$. On the other hand, if $b_a \tau_p^2 = 0$, then $v = 1$ and $\Gamma = R$.

Q.E.D.

657.05/A@57R

Proof of Proposition 3

Observe that differentiating Equation (16) with respect to ν and Γ yields precisely the same expression as Equation (A5) and, hence, the characterizations of the optimal choice of ν and Γ are the same as Equations (A2a''') and (A2b'''). These expressions are not directly affected by market risk. However, there is an indirect effect because Equation (14b) implies that q^+ is influenced by $\hat{\delta}$.

To derive the result if $\lambda_a = 0$ and the principal chooses λ_c , we first note that Equations (13a) and (13b) characterize the manager's choice of a and q given ν, Γ , and λ_c (with $\lambda_a = 0$). The manager's risk premium is:

$$\pi(\nu, \Gamma, \lambda_c) = \frac{1}{2} b_a [\nu^2 g(q^+)^2 (\tau_s^2 + \zeta \sigma_m^2) + \lambda_c^2 \sigma_m^2 + 2\nu g(q^+) \lambda_c \zeta \sigma_m^2],$$

and the principal's unconstrained objective function is:

$$\begin{aligned} \underset{\nu, \Gamma, \lambda_c}{\text{maximize}} \quad & \beta \{h(a) + \bar{\delta} g(q^+) - [c^0 + Ra^+ + \pi(\nu, \Gamma, \lambda_c)] \\ & - b_m [(1 - \nu)g(q^+)\zeta - \lambda_c] \sigma_m^2\} - q. \end{aligned}$$

The first-order condition for λ_c is $\lambda_c = \lambda^0 - \nu \zeta g(q^+)$, which, when inserted with $\lambda_a = 0$ into Equations (13a) and (13b), yields Equations (14a) and (14b), and when inserted into the first-order conditions for ν and Γ yields the same results as Equations (A2a''') and (A2b''').

If $\tau_p^2 = \tau_s^2 = 0$, then Equation (A2b''') implies $\Gamma = R$, which in turn implies $\nu = 1$.

Q.E.D.

Proof of Proposition 4

With $\lambda_c = \lambda_a = 0$, the manager's first-order conditions (see Equation [13]) and risk premium are:

$$\nu h'(a^+) = R, \quad (\text{A6a})$$

$$\bar{\delta} g'(q^+) = \Gamma + b_a \nu g'(q^+) g(q^+) \tau_q^2, \quad (\text{A6b})$$

$$\pi(\nu, \Gamma) = \frac{1}{2} b_a \nu^2 g(q^+)^2 \tau_q^2, \quad (\text{A7})$$

where $\tau_q^2 = \tau_s^2 + \zeta^2 \sigma_m^2$, and the principal's unconstrained objective function is:

$$\begin{aligned} \underset{\nu, \Gamma}{\text{maximize}} \quad & V^+(\nu, \Gamma) = \beta \{h(a) + \bar{\delta} g(q^+) - [c^0 + Ra^+ + \pi(\nu, \Gamma)] \\ & - b_m (1 - \nu) g(q^+) \zeta \sigma_m^2\} - q^+. \end{aligned} \quad (\text{A8})$$

Differentiating Equations (A7) and (A8) with respect to Γ yields:

$$\pi_\Gamma(\nu, \Gamma) = b_a \nu^2 \tau_q^2 g'(q^+) g(q^+) q_\Gamma^+, \quad (\text{A9})$$

$$V_\Gamma^+(\nu, \Gamma) = \beta \{\bar{\delta} g'(q^+) q_\Gamma^+ - \pi_\Gamma(\nu, \Gamma) - b_m (1 - \nu) g'(q^+) q_\Gamma^+ \zeta \sigma_m^2\} - q_\Gamma^+. \quad (\text{A10})$$

Setting Equation (A10) equal to 0, substituting Equations (A6b) and (A9), and rearranging terms yields:

P, 10, 644

$$\Gamma = R - (1 - \nu)g'(q^+)[b_a\nu g(q^+)\tau_q^2 - b_m\zeta\sigma_m^2]. \quad (\text{A11})$$

Differentiating Equations (A7) and (A8) with respect to ν yields:

$$\pi_\nu(\nu, \Gamma) = b_a[\nu g(q^+)^2 + \nu^2 g'(q^+)g(q^+)q_v^\dagger]\tau_q^2, \quad (\text{A12})$$

$$\begin{aligned} V_\nu^\dagger(\nu, \Gamma) = & \beta\{h'(a^+)a_v^\dagger + \bar{\delta}g'(q^+)q_v^\dagger - Ra_v^\dagger - \pi_\nu(\nu, \Gamma) \\ & + b_m[g(q^+) - (1 - \nu)g'(q^+)q_v^\dagger]\zeta\sigma_m^2\} - q_v^\dagger. \end{aligned} \quad (\text{A13})$$

Substituting Equations (A6a), (A6b), and (A12) into Equation (A13) and then substituting in Equation (A11) yields:

$$1 - \nu = \frac{[b_a\nu g(q^+)\tau_q^2 - b_m\zeta\sigma_m^2]g(q^+)}{h'(a^+)a_v^\dagger}. \quad (\text{A14})$$

Substituting Equation (A14) into Equation (A11) implies $\Gamma \leq R$, with strict inequality if $\nu \neq 1$. Observe that $\nu < (=, >) 1$ if $b_ag(q^+)\tau_q^2 > (=, <) b_m\zeta\sigma_m^2$. **Q.E.D.**

REFERENCES

- Anthony, R. N., and V. Govindarajan. 2000. *Management Control Systems*. 10th edition. New York, NY: McGraw-Hill/Irwin.
- Antle, R., and G. Eppen. 1985. Capital rationing and organizational slack in capital budgeting. *Management Science* 31: 161–174.
- , and J. Fellingham. 1990. Resource rationing and organizational slack in a two-period model. *Journal of Accounting Research* 28: 1–24.
- Atkinson, A. A., and G. A. Feltham. 1981. Information in capital markets: An agency theory perspective. Working paper, University of British Columbia, Vancouver, Canada.
- Bushman, R. M., and R. J. Indjejikian. 1993. Accounting income, stock price, and managerial compensation. *Journal of Accounting and Economics* 16: 3–23.
- , ———, and M. Penno. 2000. Private pre-decision information, performance measure congruity and the value of delegation. *Contemporary Accounting Research* 17: 561–587.
- Christensen, P. O., and G. A. Feltham. 1993. Communication in multiperiod agencies with production and financial decisions. *Contemporary Accounting Research* 9: 706–744.
- Datar, S., S. L. Kulp, and R. Lambert. 2001. Balancing performance measures. *Journal of Accounting Research* 39: 75–92.
- Demski, J. S., and R. A. Dye. 1999. Risk, return, and moral hazard. *Journal of Accounting Research* 37: 27–55.
- Dutta, S., and S. Reichelstein. 2002. Controlling investment decisions: Depreciation and capital changes. *Review of Accounting Studies* (forthcoming).
- Feltham, G. A., and J. Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review* 69: 429–453.
- , and J. A. Ohlson. 1995. Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research* 11: 689–731.
- , and ———. 1996. Uncertainty resolution and the theory of depreciation measurement. *Journal of Accounting Research* 34: 209–234.
- , and ———. 1999. Residual earnings valuation with risk and stochastic interest rates. *The Accounting Review* 74: 165–183.
- , and M. G. H. Wu. 2000. Public reports, information acquisition by investors, and management incentives. *Review of Accounting Studies* 5: 155–190.
- , and ———. 2001. Incentive efficiency of stock versus options. *Review of Accounting Studies* 6: 7–28.

- Fischer, P. E. 1999. Managing employee compensation risk. *Review of Accounting Studies* 4: 45–60.
- Hilton, R. W., M. W. Maher, and F. H. Selto. 2000. *Cost Management: Strategies for Business Decisions*. New York, NY: McGraw-Hill/Irwin.
- Horngren, C. T., G. Foster, S. M. Datar, and H. D. Teall. 2000. *Cost Accounting: A Managerial Emphasis*. Second Canadian edition. Scarborough, Ontario, Canada: Prentice Hall Canada.
- Hughes, P. 1986. Signalling by direct disclosure under asymmetric information. *Journal of Accounting and Economics* 8: 119–142.
- . 1988. Risk sharing and valuation under moral hazard. In *Economic Analysis of Information and Contracts*, edited by G. A. Feltham, A. H. Amershi, and W. T. Ziemba, 247–268. Boston, MA: Kluwer Academic Publishers.
- Indjejikian, R., and D. Nanda. 1999. Dynamic incentives and responsibility accounting. *Journal of Accounting and Economics* 27: 177–201.
- Jin, L. 2001. CEO compensation, diversification and incentives. Working paper, Massachusetts Institute of Technology, Cambridge, MA.
- Kaplan, R. S., and A. A. Atkinson. 1998. *Advanced Management Accounting*. 3rd edition. Upper Saddle River, NJ: Prentice Hall.
- Lambert, R. 2001. Contracting theory and accounting. *Journal of Accounting and Economics* 32: 3–87.
- Leland, H., and D. Pyle 1977. Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance* 32: 371–387.
- Mirrlees, J. A. 1999. The theory of moral hazard and unobservable behaviour: Part I. *Review of Economic Studies* 66: 3–21. Working paper, Nuffield College, Oxford, 1975.
- Melumad, N., and S. Reichelstein. 1987. Centralization versus delegation and the value of communication. *Journal of Accounting Research* 25 (Supplement): 1–18.
- Ohlson, J. A. 1990. A synthesis of security valuation theory and the role of dividends, cash flows, and earnings. *Contemporary Accounting Research* 6: 648–676.
- . 1995. Earnings, book values, and dividends in security valuation. *Contemporary Accounting Research* 11: 661–687.
- Rubinstein, M. 1973. A mean variance synthesis of corporate financial theory. *Journal of Finance* 28: 167–181.
- Stewart, G. B., III. 1991. *The Quest for Value*. New York, NY: Harper Collins Publishers.
- Wilson, R. 1968. The theory of syndicates. *Econometrica* 36: 119–132.

The Effect of Legal Environment on Voluntary Disclosure: Evidence from Management Earnings Forecasts Issued in U.S. and Canadian Markets

Stephen P. Baginski

John M. Hassell

Michael D. Kimbrough

Indiana University

ABSTRACT: Citing fear of legal liability as a partial explanation, prior research documents (1) managers' reluctance to voluntarily disclose management earnings forecasts, and (2) greater forecast disclosure frequencies in periods of bad news. We provide evidence on how management earnings forecast disclosure differs between the United States (U.S.) and Canada, two otherwise similar business environments with different legal regimes. Canadian securities laws and judicial interpretations create a far less litigious environment than exists in the U.S. We find a greater frequency of management earnings forecast disclosure in Canada relative to the U.S. Further, although U.S. managers are relatively more likely to issue forecasts during interim periods in which earnings decrease, Canadian managers do not exhibit that tendency. Instead, Canadian managers issue more forecasts when earnings are increasing, and their forecasts are of annual rather than interim earnings. Also consistent with a less litigious environment, Canadian managers issue more precise and longer-term forecasts. These findings hold after controlling for other determinants of management earnings forecast disclosure that might differ between

We thank workshop participants at the University of Georgia, Washington University, and Indiana University for comments on earlier versions of this paper. We also benefited from discussions with Jamie Pratt, Jerry Salamon, Jim Wahlen, Grace Pownall, Jim Seida, and discussants' comments by Paquita Davis-Friday and Lenny Soffer in conjunction with our presentations of this paper at the Second Globalization Conference (2000), jointly sponsored by the American Accounting Association and the British Accounting Association, and the 2000 American Accounting Association Annual Meeting, respectively. Professor Baginski thanks the Center for International Business Education Research (CIBER) at Indiana University for support of the study. We also gratefully acknowledge the contribution of I/B/E/S International Inc. for providing earnings per share forecast data. These data have been provided as part of a broad academic program to encourage earnings expectations research.

*Submitted August 2000
Accepted June 2001*

the two countries—firm size, earnings volatility, information asymmetry, growth, capitalization rates, and membership in high-technology and regulated industries.

Keywords: *management earnings forecasts; legal liability; voluntary disclosure.*

Data Availability: *The data used in this study are publicly available from the sources listed in the text.*

I. INTRODUCTION

This paper reports empirical evidence of differences in management earnings forecast disclosure behavior between the United States and Canada. We expect that management forecast disclosure differs between these two countries because of the substantial differences in their legal environments that affect the success of class action, securities-related litigation. Shareholders can bring lawsuits against U.S. companies under SEC Rule 10b-5. Upon observing a large stock price decline (often at an earnings announcement date), plaintiffs in these class action suits typically argue that managers failed to promptly disclose material bad news, causing plaintiffs to buy an overvalued security that declined in value after management revealed the material information.¹ Plaintiffs in the U.S. have favorably settled many of these suits out of court. In contrast, Canadian securities laws and legal interpretations make it difficult to win class action, securities-related suits, especially suits related to the secondary market for securities.

Management forecasts affect security prices (Patell 1976; Ajinkya and Gift 1984; Waymire 1984). Therefore, it is important to understand whether and how legal regimes affect forecast disclosure. On one hand, potential litigation costs associated with failure to meet a management forecast mitigate “cheap talk,” and thus may enhance the credibility of publicly disclosed forecasts. On the other hand, fear of litigation may alter the conditions under which managers issue forecasts.

As an example of how litigation fears could alter *which* forecasts are released, Skinner (1994) argues that fear of legal liability relating to Rule 10b-5 suits gives U.S. managers incentive to hasten disclosure of bad news, because (1) early disclosure is a defense against shareholders’ claims that management withheld adverse information, and (2) early disclosure shortens the period in which investors can qualify as members of a class action suit. Based on a sample of 93 NASDAQ firms, Skinner (1994) provides evidence that, because of asymmetric incentives to provide timely bad news disclosure, managers are more likely to issue imprecise interim forecasts during periods of large negative earnings surprises than during periods of large positive surprises. Similarly, in a sample of 565 NYSE and AMEX firms, Kasznik and Lev (1995) find that managers are twice as likely to disclose forecasts during periods of large negative earnings news than during periods of large positive earnings news. Policymakers also argue that the threat of litigation reduces a manager’s tendency to provide *any* voluntary disclosure of forward-looking information because of the possibility that earnings may fall short of the forecast (e.g., Breeden 1995). In fact, Congress enacted the Private Securities Litigation Reform Act of 1995 to encourage firms to disclose forward-looking information by reducing legal exposure.

¹ Technically, management’s failure to disclose material bad news that it should have released to correct misleading or inaccurate prior disclosures triggers legal liability. In practice, however, the mere existence of undisclosed bad news increases the firm’s exposure to shareholder lawsuits (see Skinner [1994] and Bamber and Cheon [1998] for related arguments).

Using a sample of U.S. and Canadian firms from 1993 to 1996, we find that Canadian firms issue more management earnings forecasts than do U.S. firms. Further, although U.S. managers are relatively more likely to issue forecasts during interim periods in which earnings decrease, Canadian managers do not exhibit that tendency. Instead, Canadian managers issue more forecasts when earnings are increasing, and their forecasts are of annual rather than interim earnings. Also consistent with a less litigious environment, Canadian managers issue longer-term and more precise forecasts. These Canadian firms' different disclosure patterns persist after controlling for other conditions that might lead to differential disclosure, including firm size, earnings volatility, information asymmetry, growth, capitalization rates, and membership in high-technology and regulated industries.

Our study contributes to management forecast research by investigating the relation between legal regime and specific management forecast disclosure properties—incidence, timeliness, and precision. Although Skinner's (1994) and Kasznik and Lev's (1995) results are consistent with legal liability leading to more disclosure in bad news periods, it is not clear that the legal environment is driving U.S. firms to issue bad news management forecasts in interim periods. First, extant empirical evidence is mixed on the effectiveness of a bad news management forecast in avoiding litigation. Francis et al. (1994) find that (1) substantial declines in earnings are not sufficient to cause litigation; (2) early disclosure of adverse earnings news is not an effective deterrent to litigation, and failure to provide early disclosure of adverse earnings news does not necessarily lead to litigation; and (3) in a litigation sample of 45 observations, 28 suits were based on a management forecast or other pre-earnings release disclosure rather than on the subsequently released actual earnings announcement. Further, Skinner (1997) finds no evidence that management forecast disclosure increases the probability of lawsuit dismissal. However, he provides some evidence that, after controlling for estimated shareholder damages, timely disclosure reduces settlement amounts. Finally, Johnson et al. (2001) interpret their finding that bad news disclosure frequency increases after the Private Securities Litigation Reform Act of 1995 as inconsistent with the argument that reduced legal exposure leads to proportionally fewer bad news disclosures.

Second, with the exception of Frost (2001) and Johnson et al. (2001), prior research has not examined management forecast disclosures in different legal environments.² Skinner (1994, 1997), Francis et al. (1994), and Kasznik and Lev (1995) examine disclosures made in the highly litigious U.S. environment. Our finding that managers are more likely to issue forecasts (both bad and good news) in the less litigious Canadian environment is consistent with Frost (2001) and Johnson et al. (2001), who conclude that managers are more likely to issue voluntary disclosures in less litigious environments. Further, our finding that firms issue proportionally fewer bad news forecasts in the Canadian legal environment (relative

² Johnson et al. (2001) capitalize on an expected change in legal liability created by passage of the Private Securities Litigation Reform Act of 1995 in the U.S. to test whether management forecasting behavior changed. They examine 523 firms in three high-technology industries and find that managers issue more good news and bad news management earnings and sales forecasts after the Act. However, confining the analysis to high-technology firms limits the results' generalizability. The extent that the Act reduced legal exposure was likely not clear to most managers and their legal counsels. Although the rate of post-Act fraud suit dismissals increased from 40 to 60 percent (Levine and Pritchard 1998), 40 percent of the suits continued. Further, Grundfest and Perino (1997) document a shift of post-Act filings from federal to state courts, presumably to avoid the defendant's additional protections under the Federal Act.

Frost (2001) provides comparative descriptive evidence on disclosure of forward-looking information in the U.S. and several European countries for firms in the manufacturing sector. Samples of 80 firms in each country reveal that managers issue more forecasts and more precise forecasts in the non-U.S. countries where legal and regulatory environments are less stringent.

to good news forecasts) is consistent with Skinner's (1994) and Kasznik and Lev's (1995) conclusions that legal liability increases U.S. firms' incentives to disclose management earnings forecasts when earnings news is bad. However, our evidence is inconsistent with Johnson et al.'s (2001) evidence that U.S. high-tech firms do not issue proportionally fewer bad news forecasts after passage of the 1995 Private Securities Litigation Reform Act (which was intended to reduce legal liability).

The results of our study are also relevant to accounting policymakers and securities regulators who believe litigation reform can improve financial reporting. The AICPA's Jenkins Committee identified fear of litigation as an obstacle to providing forward-looking information (AICPA 1994). Managers and auditors complained about excessive and frivolous litigation; accordingly, Congress provided additional safe harbors in the Private Securities Litigation Reform Act of 1995 for parties associated with good-faith projections. Evidence on the effect of legal liability on management forecasting can help policymakers understand the costs of forecast disclosure, which is important if they expect cost-reduction strategies to create incentives for forecast disclosure.

In Section II, we develop hypotheses about how differences in legal environment are likely to affect Canadian and U.S. management forecasting. We describe the sample in Section III and the research design in Section IV. We present the primary results in Section V and the results of supplemental tests in Section VI. In Section VII, we discuss the study's limitations and provide suggestions for future research.

II. MANAGEMENT FORECAST DISCLOSURES IN THE U.S. AND CANADA

Evidence suggests that U.S. management forecasting behavior has changed over time. Patell (1976) suggested that, in the 1960s and 1970s, managers selectively released good news. In the early 1980s, Ajinkya and Gift (1984) and Waymire (1984) documented symmetric forecasting. Then, in the late 1980s and early 1990s, litigation against companies and their auditors increased. Accordingly, Skinner (1994) investigated whether legal liability creates an asymmetric incentive favoring timely management forecasting of bad news. Using a sample of U.S. management forecasts from the 1980s, his results are largely consistent with legal liability-based arguments: (1) management issues more forecasts in periods of bad earnings news than in periods of good news (evidence of asymmetric incentives); (2) bad news interim forecasts occur in imprecise form (e.g., range rather than point, open-interval rather than closed-interval); and (3) the bad news disclosures occur in interim periods. Although these findings are clearly consistent with legal-liability arguments, Skinner (1994) does not vary legal regime.

Differences between the U.S. and Canadian legal systems create a natural experiment with two different legal regimes. Clarkson and Simunic (1994) note that, unlike the U.S., (1) courts in Canada generally require unsuccessful plaintiffs to pay the costs of a successful defendant; (2) because plaintiffs have no absolute right to a jury trial in Canada, judges typically hear technical cases and are less likely to award large settlements based on emotional plaintiff appeals; (3) Canadian discovery rules limit whom the plaintiff can examine under oath; (4) Canadian provinces generally do not allow lawyers to work on a contingent-fee basis, and when allowed, the practice is not widespread; and (5) it is considerably more difficult to bring a class action suit in Canada. Clarkson and Simunic (1994) argue that auditors and managers operating in Canada are therefore likely to make decisions with the expectation of lower litigation costs than would be the case in the United States. Although McConomy (1998) notes that the Ontario Class Proceedings Act (effective in 1993) made class action lawsuits more likely, Paskell-Mede (1994, 1999) and Grossman (1996) point out that class action suits remain rare in Canada and continue to face such significant impediments that Canada is still a far less litigious environment than is the U.S.

A particularly important difference between Canadian and U.S. securities legislation is the notion of “fraud on the market” and its relation to secondary security markets (Paskell-Mede 1999). In common law actions, a plaintiff must prove that (1) a misrepresentation exists, (2) the misrepresentation was negligent, and (3) the plaintiff suffered a loss due to reliance on the misrepresentation. In U.S. courts, the “fraud on the market” concept provides a proof that the plaintiff relied on share price that efficiently reflected the misrepresentation. Judges in Canadian courts refuse to apply the “fraud on the market” concept, stating that it has no place in the Canadian legal system. Instead, Canadian law places the burden of proof on the defendant to prove that the plaintiff did not rely on the misrepresentation, but only in the case of initial security offerings. That is, once a plaintiff proves that management made a negligent misrepresentation in the prospectus, the law deems the plaintiff’s reliance to have occurred unless the defendant can prove that the plaintiff did not rely on the misrepresentation when purchasing shares (often by proving that the plaintiff was aware of the misrepresentation). However, in the case of secondary trading, the burden of proving reliance remains with the plaintiff. Because plaintiffs in Canadian secondary-markets-related suits must prove individual reliance on a misrepresentation, but cannot use the “fraud on the market” argument, they face a significant barrier to successful litigation.³

Hypothesis Development

We test two general hypotheses. First, the concern that legal liability costs curb public disclosure of either favorable or adverse forward-looking information lies at the core of U.S. policymakers’ arguments for increased forward-looking information disclosure. In addition, under economic models of symmetric belief adjustment (King et al. 1990), management forecasts increase the firm’s exposure to legal liability. Our first hypothesis (stated in the alternative form) is:

H1: Canadian firms, faced with a less-litigious legal environment, engage in more voluntary information disclosure relative to U.S. firms (*degree of information disclosure hypothesis*).

We define “more voluntary information disclosure” to include: (1) a higher forecast frequency, (2) more precise forecasts, and (3) forecasts over longer horizons. Accordingly, we perform three sets of tests of H1 that correspond to these three measures of information disclosure. Hypothesis 1 predicts that (1) Canadian firms are more likely to disclose management earnings forecasts because managers do not fear the legal liability associated with being inaccurate; (2) Canadian managers disclose more precise forecasts even though inaccuracy is easier to detect *ex post* in more precise forecasts; and (3) Canadian managers disclose forecasts over longer horizons even though the likelihood of inaccuracy increases in forecast horizon.

Our second hypothesis (also stated in the alternative form) relates to prior evidence suggesting that litigation exposure prompts U.S. firms to strategically base the frequency and properties of management forecasts on whether earnings are increasing or decreasing. Specifically, when earnings are decreasing (i.e., in bad news periods), U.S. firms tend to be: (1) more likely to issue a forecast; (2) more likely to issue the forecast in an imprecise form; and (3) more likely to issue the forecast over a short-horizon period. We expect

³ Empirical evidence suggests that even the stronger Canadian laws for primary offerings do not introduce significant fear of litigation. Clarkson and Simunic (1994) note that approximately 50 percent of Canadian initial public offerings include a management forecast in the offering prospectus (a virtually nonexistent practice in U.S. initial public offerings), and these management forecasts are on average optimistic (Clarkson et al. 1992).

Canadian firms' forecast disclosure properties (frequency, precision, and horizon) to be less associated with earnings decreases:

H2: Canadian firms, faced with a less-litigious legal environment, engage in voluntary information disclosure behaviors that are less related to earnings decreases than do U.S. firms (*bad-news-related behavior hypothesis*).

Because prior research has shown that U.S. firms' forecast frequency, precision, and horizon are related to earnings decreases, H2 is rejected if either Canadian forecast frequency, precision, and horizon are unrelated to the sign of earnings changes or if these Canadian forecast disclosure properties are related to earnings *increases*. In either case, Canadian firms do not engage in the kinds of forecasting behaviors that are related to fear of legal liability resulting from bad earnings news.

Country Similarities that Strengthen Internal Validity

The similar business and investment environments in the U.S. and Canada support our inferences that differences across the two countries in the frequency, precision, and horizon of management forecasts are attributable to differences in legal regimes. For example, researchers often group together U.S. and Canadian firms in international tests of value relevance of accounting data, primarily because these firms are similar on so many dimensions and because empirical evidence shows that the association between accounting data and stock prices is relatively similar in the two countries (Ali and Hwang 2000). The U.S. and Canada share important determinants of the value relevance of accounting data—a market-oriented financial system, private sector standard setting, low influence of tax rules on financial accounting measurements, and higher spending on external auditing (Ali and Hwang 2000). Although generally accepted accounting principles differ between Canada and the U.S., Bandyopadhyay et al. (1994) and Barth and Clinch (1996) find that adjustments from Canadian to U.S. GAAP are irrelevant for valuing cross-listed Canadian companies.

Investor sophistication and market efficiency with respect to the interpretation of forecasted data also appear similar in the two countries. McNichols (1989) provides empirical evidence that U.S. investors can assess the *ex ante* firm-specific bias in seasoned firms' management forecasts. Clarkson et al. (1992) find similar investor sophistication for Canadian firms' management forecasts issued in initial public offerings. With respect to market efficiency, Freund et al. (1997) conclude that the Toronto Stock Exchange (TSE) is informationally efficient. Brown et al. (1991) also document that investors privy to financial analyst forecasts could generate abnormal returns on the TSE, but that the returns are not significantly higher than transactions costs if the trading strategy is not implemented until after the forecasts become public (also see L'Her and Suret [1991] for similar findings).⁴ In summary, empirical evidence suggests that revelation of Canadian managers' private

⁴ Our data requirements result in a Canadian management forecast sample that is almost entirely composed of large firms listed on the TSE, approximately three-quarters of which have analyst following. The market for shares of large firms with analyst following is more likely to be informationally efficient. Further, if managers release forecasts to move stock prices, a lack of market efficiency would likely result in fewer Canadian management forecasts relative to the U.S. (a bias against the null hypothesis that we reject in tests described later). Also, a bad news management forecast would be a more beneficial legal liability-limiting strategy in Canada because the company could reap the benefits of public disclosure without the associated negative stock price reaction (a bias in favor of the null hypothesis that Canadian management forecasts exhibit the same bad news bias as do U.S. management forecasts—a null that we reject in tests described later).

information in management forecast form should lead to a relatively complete and immediate price reaction.⁵

Canadian and U.S. exchanges have similar rules on material news disclosure. The TSE's "Policy Statement on Timely Disclosure and Related Guidelines" requires firms to provide investors with equal access to all material information. Material information is "any information relating to the business and affairs of a company that results in or would reasonably be expected to result in a significant change in the market price or value of any of the company's listed securities" (TSE 2000, 1). The TSE requires "immediate disclosure" of "firm evidence of significant increases or decreases in near-term earnings prospects," although "forecasts of earnings and other financial forecasts need not be disclosed" (TSE 2000, 4). Further, the TSE encourages wide dissemination through news services (specifically mentioning Dow Jones). Exceptions to these rules allow delay for proprietary information. Guidelines and rules of law also prohibit insider trading.

Finally, differences in earnings management opportunities may exist between the two countries. However, to create an alternative explanation for our results, (1) differences in earnings management opportunities must exist between the two countries; (2) earnings management must be a successful means of limiting legal liability; and (3) when considering the menu of alternative actions to limit legal liability (e.g., engage in earnings management, issue a voluntary management forecast disclosure of bad news, do not provide a forecast that could turn out to be wrong, etc.), the action chosen must differ between the U.S. and Canada because one alternative performs better on a cost/benefit trade-off. If these conditions all hold, then a finding of, say, fewer management forecasts in the U.S. might indicate that U.S. managers can more easily manage subsequently revealed earnings, that such earnings management fools investors about firm performance, and that earnings management is more beneficial to the firm than voluntary disclosure. We are unaware of any empirical evidence on differential benefits to earnings management between the U.S. and Canada, but we cannot rule out this possibility because controlling for earnings management differences across countries is beyond the scope of our study.

In summary, many similarities between U.S. and Canadian environments strengthen the internal validity of our comparative analysis. Still, differences exist that may affect management forecasts—information asymmetry, analyst coverage, earnings volatility, firm size, membership in high-tech and regulated industries, growth, and capitalization rates. Our empirical tests control for these conditions.

III. SAMPLE

Periods Available for Forecasting

Our sample includes all quarterly and annual periods of U.S. and Canadian firms with available data on Compustat, including beginning-of-period price data, shares, earnings per share, and adjustment factors for the 1993–1996 period.⁶ Table 1, Panel A reports that

⁵ In an untabulated regression of percentage change in analyst forecasts on percentage change in expectations conveyed by management forecasts, the slope coefficient is significantly positive ($p < 0.001$). Significant differences do not exist (1) between the slope coefficient for Canadian and U.S. firms, and (2) for the Canadian firms, between locally listed firms and firms cross-listed in the U.S. These results suggest that analysts update their expectations in response to management forecasts.

⁶ Like most research, our sample is limited to firms covered by Compustat, which covers only about half of the firms reporting to the SEC in the middle of our sample period. (In the 1994 manual, Compustat reports coverage of approximately 6,800 U.S. companies; Guenther and Rosman [1994] report that 13,424 firms filed with the SEC in 1994.) Although we do not have data on the number of Canadian firms in 1994, the most recent TSE web site lists 1,407 traded firms. The 1998 Compustat tape includes 674 TSE firms, nearly half of the TSE population, which according to its web site, represents 95 percent of market capitalization in Canada. Thus, Compustat Canadian firm coverage spans an important segment of the Canadian market.

TABLE 1
Sample Selection Criteria and Sample Characteristics

Panel A: Selection of 1993–1996 Firm-Quarters and Firm-Years

	<i>U.S. Firms</i>	<i>Canadian Firms</i>	<i>Total</i>
Firm-years available on Compustat	51,434	5,073	56,507
Missing data	(24,895)	(2,102)	(26,997)
Surviving firm-years	<u>26,539</u>	<u>2,971</u>	<u>29,510</u>
Firm-quarters available on Compustat	132,737	19,694	152,431
Missing data	(52,600)	(13,590)	(66,190)
Surviving firm-quarters	<u>80,137</u>	<u>6,104</u>	<u>86,241</u>
Total firm-years and firm-quarters sample	<u>106,676</u>	<u>9,075</u>	<u>115,751</u>
<i>Distributed across Following Exchanges</i>		<i>Firms</i>	<i>% of total</i>
NASDAQ		39,591	34.2
NYSE		34,412	29.7
OTC		25,583	22.2
American Stock Exchange		8,737	7.5
Toronto Stock Exchange		5,757	5.0
Regional exchanges		893	0.8
Vancouver Stock Exchange		378	0.3
Nonlisted Canadian		253	0.2
Montreal Exchange		147	0.1
Total firm-years and firm-quarters sample		<u>115,751</u>	<u>100.0</u>

Panel B: Selection of 1993–1996 Management Earnings Forecasts

	<i>Issued by U.S. Firms</i>	<i>Issued by Canadian Firms</i>	<i>Total</i>
Earnings forecasts identified by keyword search of Dow Jones News Retrieval Service	1,688	391	2,079
Lost due to changes in Compustat coverage	(281)	(78)	(359)
Listed on Compustat	<u>1,407</u>	<u>313</u>	<u>1,720</u>
Missing Compustat data	(74)	(36)	(110)
Complete data on Compustat	1,333	277	1,610
Forecasts of 1992 and 1997 results	(112)	(26)	(138)
Long-range (> 365 days) forecasts	(57)	(32)	(89)
Final management earnings forecast sample	<u>1,164</u>	<u>219</u>	<u>1,383</u>
<i>Distributed across Following Exchanges</i>		<i>Forecasts</i>	<i>% of Total</i>
NYSE		649	46.9
NASDAQ		376	27.2
Toronto Stock Exchange		206	14.9
Other U.S.		139	10.1
Other Canadian		13	0.9
Final management earnings forecast sample		<u>1,383</u>	<u>100.0</u>

(Continued on next page)

TABLE 1 (Continued)

<i>Distributed across Firms</i>	<i>Firms (%)</i>	<i>Forecasts (%)</i>
U.S. firms that issued:		
one forecast	398 (57.9)	398 (34.2)
two forecasts	181 (26.4)	362 (31.1)
three forecasts	64 (9.3)	192 (16.5)
four or more forecasts	44 (6.4)	212 (18.2)
	<u>687 (100.0)</u>	<u>1,164 (100.0)</u>
Canadian locally listed firms that issued:		
one forecast	38 (59.4)	38 (34.9)
two forecasts	16 (25.0)	32 (29.3)
three forecasts	5 (7.8)	15 (13.8)
four or more forecasts	5 (7.8)	24 (22.0)
	<u>64 (100.0)</u>	<u>109 (100.0)</u>
Canadian cross-listed firms that issued:		
one forecast	20 (40.0)	20 (18.2)
two forecasts	16 (32.0)	32 (29.1)
three forecasts	8 (16.0)	24 (21.8)
four or more forecasts	6 (12.0)	34 (30.9)
	<u>50 (100.0)</u>	<u>110 (100.0)</u>

56,507 firm-years and 152,431 firm-quarters are available. Periods with complete data totaled 115,751. NASDAQ was the most represented U.S. exchange (34.2 percent of all periods in the two-country combined sample), and the TSE was the most represented Canadian exchange (5.0 percent of all periods in the combined sample). In contrast to most prior legal-liability-related research and in order to enhance our results' generalizability, we do not limit our tests to periods with "large" earnings surprises.

Management Earnings Forecasts

We collected management forecasts of both interim and annual earnings from the Dow Jones News Retrieval Service (DJNRS) during the entire 1993–1996 period.⁷ The DJNRS contains articles published in the *Wall Street Journal* and *Barron's*, as well as unpublished announcements appearing on the Broad Tape. The DJNRS covers the major news wires and newspapers in Canada, especially those specializing in corporate news. In addition, the TSE lists Dow Jones as a preferred means of disclosure. Therefore, we consider DJNRS a reasonable (and unbiased) source of management earnings forecasts. However, if DJNRS provides incomplete Canadian coverage, either because of the DJNRS physical location or because U.S. investors are less interested in smaller foreign firms, then the incomplete coverage would bias against our finding of greater forecast frequency for Canadian firms. We also do not limit the coverage period to only the fourth quarter because Canadian firms are required to file only the first three quarters of interim results (Frost and Kinney 1996).

⁷ Our DJNRS search keywords are: "expects earnings," "expects net," "expects income," "expects losses," "expects profits," "expects results," and three similar lists with first words "forecasts," "predicts," and "sees."

Table 1, Panel B presents information about the management forecast selection. Our original search identified 2,079 forecasts. We lost 359 forecasts due to changes in Compustat coverage.⁸ After deleting forecasts for periods with missing Compustat data (110), forecasts of 1992 and 1997 results (138), and long-range forecasts (89), a total of 1,383 management forecasts remain in the sample: 1,164 issued by U.S. firms and 219 issued by Canadian firms.⁹

The 1,164 U.S. forecasts are made by 687 firms (1.69 per firm). The 219 Canadian management forecasts are made by 64 Canadian firms listed only in Canada and 50 Canadian firms cross-listed on U.S. exchanges (1.92 per firm). Our primary tests do not differentiate locally listed from cross-listed Canadian firms. Pooling the locally listed and cross-listed firms reduces the power of our tests to the extent that cross-listed Canadian firms face some legal exposure in U.S. markets. However, Frost and Pownall (1994) were unable to identify a single instance of formal investigation, administrative proceeding, or court action against a foreign issuer for violation of U.S. disclosure rules. Legal analysis (Paskell-Mede 1999) also raises doubts about the extent to which cross-listed firms are exposed to liability under the U.S. legal system. Frost and Pownall (1994) cite political issues, difficulty in monitoring, and lack of cooperative agreements with other countries as barriers to enforcement of foreign-issuer compliance. Finally, Core (1997) finds that although cross-listing is associated with the magnitude of director and officer liability coverage for Canadian companies that purchase insurance, cross-listing does not explain Canadian firms' decisions to purchase the policy. In summary, cross-listing is unlikely to increase legal exposure to the level faced by U.S. firms. Accordingly, Table 1, Panel B provides descriptive data on the disclosures of the Canadian cross-listed firms vis-à-vis locally listed Canadian and U.S. firms, but our primary tests pool together the cross-listed and locally listed Canadian firms.

IV. RESEARCH DESIGN

In this section, we present our research design for tests of H1 and H2 using the three measures of information disclosure: forecast frequency, forecast precision, and length of forecast horizon.

Forecast Frequency

To test H1 and H2, we estimate the following logistic regression model for a combined sample of U.S. and Canadian companies across all potential forecasting periods ($n = 115,751$):¹⁰

$$\begin{aligned} \text{FORECAST}_{i,t} = & a_0 + a_1 \Delta \text{SIGN}_{i,t} + a_2 \text{CANADA}_{i,t} + a_3 \text{CANADA}_{i,t} \\ & \times \Delta \text{SIGN}_{i,t} + a_4 |\Delta \text{EARNINGS}_{i,t}| + a_5 \text{LSIZE}_{i,t} \\ & + a_6 \text{REGULATE}_{i,t} + a_7 \text{HIGHTECH}_{i,t} + \varepsilon_{i,t} \end{aligned} \quad (1)$$

⁸ A major source of data loss is a combination of CUSIP and name changes that we are unable to track. Conversations with Compustat personnel have not enabled us to recover these observations. Because the rate of management earnings forecast observation loss is higher for Canadian firms ($78 \div 391 = 20$ percent) than U.S. firms ($281 \div 1,688 = 16.6$ percent), and because we find greater forecast frequency for Canadian firms, the loss of observations does not affect our inferences for H1.

⁹ Canadian forecasts represent 18.8 percent of the original sample (391 of 2,079). However, 36 percent of the discarded long-range forecasts (32 of 89) are Canadian, indicating that Canadian firms provide more long-range (> 365 days) forecasts.

¹⁰ We also ran OLS regressions, and the results are qualitatively identical. We code single news stories containing both annual and interim earnings forecasts as two separate observations.

where:

$FORECAST_{i,t} = 1$ if the firm issued a management earnings forecast during the period, and 0 otherwise;

$\Delta SIGN_{i,t} =$ the sign of the earnings change, defined as 1 for $\Delta EARNINGS_{i,t} \geq 0$ (i.e., good news) and 0 for $\Delta EARNINGS_{i,t} < 0$ (i.e., bad news); where $\Delta EARNINGS_{i,t} =$ the change in earnings, defined as $(EPS_{i,t} - EPS_{i,t-k})/PRICE_{i,t-k}$, where $EPS_{i,t} =$ earnings (net income) per share for firm i in period t ; $EPS_{i,t-k} =$ earnings per share for firm i in period $t - 1$ for annual and $t - 4$ for quarterly periods; and $PRICE_{i,t-k} =$ security price for firm i at the end of period $t - 1$ for annual and $t - 4$ for quarterly periods (all obtained from Compustat);¹¹

$CANADA_{i,t} = 1$ if the potential forecasting period relates to a Canadian firm (regardless of whether the firm issued a forecast in that period), and 0 otherwise (i.e., 0 when the potential forecasting period relates to a U.S. firm);

$|\Delta EARNINGS_{i,t}| =$ the absolute value of $\Delta EARNINGS_{i,t}$;

$LSIZE =$ the log of market value of equity at the beginning of the period;

$REGULATE = 1$ if the firm is in a regulated industry, and 0 otherwise; and

$HIGHTECH = 1$ if the firm is in a high-technology industry, and 0 otherwise.

Figure 1 maps the coefficients in Equation (1) to H1 and H2. Column (1) lists the coefficient sums in earnings increase periods (i.e., good news; $\Delta SIGN = 1$) for Canadian firms (row 1) and for U.S. firms (row 2). Column (2) provides analogous coefficients for earnings decrease periods (i.e., bad news; $\Delta SIGN = 0$). The last row indicates the difference *between countries* in the propensity to issue forecasts in periods of good news ($a_2 + a_3$) and bad news (a_2). Hypothesis 1 predicts that Canadian firms issue more forecasts; thus, we expect both sets of coefficients to be positive ($a_2 + a_3 > 0$, $a_2 > 0$).

The last column of Figure 1 provides coefficients associated with differences between good and bad news periods (i.e., "sign-related" forecasting behavior) in Canada (row 1) and in the U.S. (row 2). The last row in the column shows that the coefficient a_3 measures the difference *between countries* in sign-related forecasting behavior. If legal-liability-created asymmetric forecast disclosure incentives in the U.S. lead to more forecasts in bad news periods, then we expect $a_1 < 0$. Hypothesis 2 predicts $a_3 > 0$, indicating that Canadian managers are less likely to skew forecast disclosure toward bad news periods. Given our focus on the differences associated with different legal regimes, we base all hypothesis tests on the difference between countries, not the conditions within a country.

The remaining variables in Equation (1) control for other potential nonlegal-liability-related differences between the two countries that could affect management forecasting behavior. Firms with more volatile earnings face greater risk of inaccurate forecasts, with

¹¹ Kasznik and Lev (1995) employ alternative quarterly expectation models (analyst forecasts and seasonal random-walk) and document no difference in results. To maximize sample size, results tabulated for Equation (1) are based on the random-walk (seasonal random-walk) expectations to compute annual (quarterly) earnings surprise. Replication with a random-walk for quarterly earnings does not affect our conclusions. Analyst forecast coverage is significantly greater for our U.S. firms (mean analyst coverage of 1.84) than for our Canadian firms (mean analyst coverage of 1.29, Wilcoxon rank sum test p -value < 0.001 on the difference). We were able to obtain 1/B/E/S median financial analyst forecasts for 40 percent of the sample. Replication substituting 1/B/E/S analyst forecasts when available and limiting the analysis to firms covered by analysts does not affect our inferences.

FIGURE 1
Mapping Equation (1) Coefficients into Hypothesis Tests^a

	<i>Good News Period</i> $\Delta\text{ESIGN} = 1$	<i>Bad News Period</i> $\Delta\text{ESIGN} = 0$	<i>Difference across Sign of Earnings Change</i> (column 1 – column 2)
Canadian (CANADA = 1)	$a_0 + a_1 + a_2 + a_3$	$a_0 + a_2$	$a_1 + a_3$
U.S. (CANADA = 0)	$a_0 + a_1$	a_0	a_1
Difference between countries (row 1 – row 2)	(H1 test during good news period): $a_2 + a_3 > 0$	(H1 test during bad news period): $a_2 > 0$	Difference between countries in sign-related behavior (H2): $a_3 > 0$

^a Equation (1): $\text{FORECAST}_{i,t} = a_0 + a_1 \Delta\text{ESIGN}_{i,t} + a_2 \text{CANADA}_{i,t} + a_3 \text{CANADA}_{i,t} \times \Delta\text{ESIGN}_{i,t} + a_4 |\Delta\text{EARNINGS}_{i,t}| + a_5 \text{LSIZE}_{i,t} + a_6 \text{REGULATE}_{i,t} + a_7 \text{HIGHTECH}_{i,t} + \varepsilon_{i,t}$

To map equations (2), (3), and (4) into the hypotheses, replace the “a” coefficients with “b,” “c,” and “c” coefficients, respectively, and reverse the sign predictions on b_3 and c_3 . Then refer to the following models:

Equation (2): $\text{PRECISE}_i = b_0 + b_1 \Delta\text{ESIGN}_i + b_2 \text{CANADA}_i + b_3 \text{CANADA}_i \times \Delta\text{ESIGN}_i + b_4 \text{LSIZE}_i + b_5 \text{REGULATE}_i + b_6 \text{HIGHTECH}_i + b_7 \text{FHORIZON}_i + v_i$

Equation (3): $\text{FHORIZON}_i = c_0 + c_1 \text{NEWS}_i + c_2 \text{CANADA}_i + c_3 \text{CANADA}_i \times \text{NEWS}_i + b_4 \text{LSIZE}_i + b_5 \text{REGULATE}_i + b_6 \text{HIGHTECH}_i + \mu_i$

Equation (4): $\text{ALTFHORIZON}_i = c_0 + c_1 \text{NEWS}_i + c_2 \text{CANADA}_i + c_3 \text{CANADA}_i \times \text{NEWS}_i + c_4 \text{LSIZE}_i + c_5 \text{REGULATE}_i + c_6 \text{HIGHTECH}_i + \eta_i$

$\Delta\text{ESIGN} = 1$ for $\Delta\text{EARNINGS} \geq 0$ (i.e., good news), 0 for $\Delta\text{EARNINGS} < 0$ (i.e., bad news), where $\Delta\text{EARNINGS}_{i,t} = [\text{EPS}_{i,t} - \text{EPS}_{i,t-k}] / \text{PRICE}_{i,t-k}$;

$\text{EPS}_{i,t}$ = earnings per share for firm i in period t ;

$\text{EPS}_{i,t-k}$ = earnings per share for firm i in period $t-k$;

$\text{PRICE}_{i,t-k}$ = security price for firm i in period $t-k$; and

$k = 1$ and 4 for annual and quarterly observations, respectively.

$\text{FORECAST} = 1$ if the firm issued a management forecast during the period, and 0 otherwise;

$\text{PRECISE} = 3$ (2, 1, 0) for point (range, open-interval, general impression) forecasts;

$\text{CANADA} = 1$ if the potential forecasting period relates to a Canadian firm, and 0 if it relates to a U.S. firm;

$\text{NEWS} = 1$ if the forecast is good news, and 0 if the forecast is bad news. *Good news* disclosures include management forecast point estimates above expected earnings, range estimates where the midpoint exceeds expected earnings, and all minimum estimates. Expected earnings equals the median I/B/E/S analyst forecast when available. Otherwise, expected earnings equals the random-walk EPS for annual forecasts and seasonal random-walk for quarterly forecasts. *Bad news* disclosures include point estimates below expected earnings, range estimates where the midpoint is less than expected earnings, and all maximum estimates;

FHORIZON = the number of calendar days between the date of forecast and the end of the period being forecast;

ALTFHORIZON = the number of calendar days between the date of forecast and the end of the firm's fiscal year;

LSIZE = natural log of the market value of equity at the beginning of the period;

$\text{HIGHTECH} = 1$ if a firm belongs to Pharmaceuticals (Compustat SIC codes 2833–2836), R&D Services (8731–8734), Programming (7371–7379), Computers (3570–3577), or Electronics (3600–3674) industries, and 0 otherwise; and

$\text{REGULATE} = 1$ if a firm belongs to Telephone (4812–4813), TV (4833), Cable (4841), Communications (4811–4899), Gas (4922–4924), Electricity (4931), Water (4941), or Financial (6021–6023, 6035–6036, 6141, 6311, 6321, 6331) industries, and 0 otherwise.

the associated legal and reputational costs. We control for volatility using the absolute value of the current-period earnings change, $|\Delta \text{EARNINGS}|$, because this captures current-period volatility and does not introduce the survivorship bias associated with an historical measure. We expect $a_4 < 0$ because managers are less willing to forecast when earnings are volatile. Prior research finds that larger firms issue more management forecasts (Cox 1985; Waymire 1985), so we control for the log of market value of equity at the beginning of the period, LSIZE , and expect $a_5 > 0$. Regulatory bodies require regulated firms to produce large amounts of information, and thus, management earnings forecasts are likely less beneficial (Kasznik and Lev 1995). Using Kasznik and Lev's (1995) classification scheme, REGULATE equals 1 when a firm belongs to Telephone (4812–4813), TV (4833), Cable (4841), Communications (4811–4899), Gas (4922–4924), Electricity (4931), Water (4941), or Financial (6021–6023, 6035–6036, 6141, 6311, 6321, 6331) industries, and 0 otherwise. We expect $a_6 < 0$ if regulated firms issue fewer forecasts. High-tech firms are riskier, and the greater value of private information may lead less-informed investors to demand more voluntary, publicly released management forecasts to level the informational playing field. On the other hand, high-tech firms may be less likely to issue a management forecast to avoid conveying proprietary information. Accordingly, we cannot predict the sign of a_7 . Following Kasznik and Lev (1995), HIGHTECH equals 1 if a firm belongs to Pharmaceuticals (Compustat SIC codes 2833–2836), R&D Services (8731–8734), Programming (7371–7379), Computers (3570–3577), or Electronics (3600–3674) industries, and 0 otherwise.

Forecast Precision

King et al. (1990) argue that managers are concerned about potential litigation if a forecast turns out to be inaccurate. Less precise forecasts (e.g., a range, minimum, maximum, or general impression) are less likely to turn out to be inaccurate than are point forecasts. Accordingly, researchers have argued that, when faced with perceived higher expected litigation costs, managers will issue less precise forecasts. Empirical evidence in U.S. markets is consistent with the assertion that forecasts are less precise when the firm is performing poorly (Skinner 1994; Baginski and Hassell 1997; Bamber and Cheon 1998).

Although Canadian firm managers are subject to adverse reputation effects of inaccuracy, the Canadian legal system exacts lower legal penalties for inaccuracy than does the U.S. system. Canadian managers are therefore likely to issue more precise management forecasts (H1) and to make forecast precision choices that are less likely to depend on whether the firm is performing poorly during the period (H2). To test these hypotheses, we estimate the following ordered logistic regression model for a pooled sample of all *forecasts* issued by U.S. and Canadian firms ($n = 1,383$):

$$\begin{aligned} \text{PRECISE}_i = & b_0 + b_1 \Delta \text{SIGN}_i + b_2 \text{CANADA}_i + b_3 \text{CANADA}_i \times \Delta \text{SIGN}_i \\ & + b_4 \text{LSIZE}_i + b_5 \text{REGULATE}_i + b_6 \text{HIGHTECH}_i \\ & + b_7 \text{FHORIZON}_i + v_i. \end{aligned} \quad (2)$$

We measure management forecast precision using an ordinal coding scheme that assigns the highest value to the most precise forecasts. PRECISE equals 3, 2, 1, and 0 for point, closed-interval, open-interval, and general impression forecasts, respectively. Hypothesis 1 predicts that Canadian firms will issue more precise forecasts because the legal penalties for inaccuracy are smaller. For earnings decreases, this suggests that $b_2 > 0$, and for earnings increases, it suggests that $b_2 + b_3 > 0$. If fear of legal liability leads U.S. firms to

issue less precise forecasts when the firm is performing poorly, then $b_1 > 0$. Hypothesis 2 predicts that Canadian forecast precision is less skewed toward poor performance than is U.S. forecast precision ($b_3 < 0$).¹²

Based on Baginski and Hassell's (1997) conclusion that larger firms garner lower benefits of precise forecasting, we expect a negative coefficient on firm size ($b_4 < 0$). As in Baginski and Hassell (1997) and Bamber and Cheon (1998), we also control for forecast horizon (FHORIZON), the number of calendar days between the forecast and period-end. The longer the forecast horizon, the greater management's uncertainty about earnings, which should lead to less precise forecasts. Thus, we expect $b_7 < 0$. We continue to expect $b_5 < 0$ (less precise forecasts for regulated firms), and we make no prediction for b_6 .

Length of Forecast Horizon

Longer forecast horizons provide investors with information on a timelier basis. However, longer-horizon forecasts have a greater chance of being *ex post* inaccurate and, as a result, increase the firm's legal exposure. Canadian firms are more likely to enjoy the benefits of providing investors with more timely information because Canadian firms face lower legal exposure when issuing longer-horizon forecasts. Thus, H1 predicts that Canadian firms issue longer-horizon forecasts for both good and bad news.

Skinner (1994) finds that, faced with substantial legal exposure, U.S. firms' longer-horizon annual forecasts reflect primarily good news, but their shorter-horizon interim forecasts reflect primarily bad news. Hypothesis 2 predicts that Canadian firms, facing relatively less legal exposure, are less likely to restrict bad news forecasts to short horizons.

To test these predictions, we estimate the following OLS regression models for the pooled sample of U.S. and Canadian management forecasts ($n = 1,204$):

$$\begin{aligned} \text{Model 1: FHORIZON}_i = & c_0 + c_1 \text{NEWS}_i + c_2 \text{CANADA}_i + c_3 \text{CANADA}_i \\ & \times \text{NEWS}_i + c_4 \text{LSIZE}_i + c_5 \text{REGULATE}_i \\ & + c_6 \text{HIGHTECH}_i + \mu_i. \end{aligned} \quad (3)$$

$$\begin{aligned} \text{Model 2: ALTFHORIZON}_i = & c_0 + c_1 \text{NEWS}_i + c_2 \text{CANADA}_i + c_3 \text{CANADA}_i \\ & \times \text{NEWS}_i + c_4 \text{LSIZE}_i + c_5 \text{REGULATE}_i \\ & + c_6 \text{HIGHTECH}_i + \eta_i. \end{aligned} \quad (4)$$

As in Baginski and Hassell (1997) and Skinner (1997), we operationalize forecast horizon (FHORIZON) as the number of calendar days until period-end, regardless of whether the period is an interim or annual forecasting period. However, viewing an interim forecast as part of an annual period suggests an alternative measure of horizon (ALTFHORIZON), the number of days until fiscal year-end. For example, a first-quarter forecast 15 days before quarter-end is 288 days before year-end. Accordingly, we estimate the model for each of the two forecast horizon definitions.

¹² An alternative contemporaneous measure of earnings change directly compares the management forecast to prevailing analyst expectations. However, this measure is problematic because its construction is a function of forecast precision, the dependent variable. Therefore, we use the period's earnings change in the precision tests, consistent with Bamber and Cheon (1998), who use earnings change rather than forecast-based news in their model to explain forecast precision.

When investigating the relation between legal liability and forecast horizon, Skinner (1994) defines good and bad news in terms of the change in contemporaneous expectations. Accordingly, we define a dichotomous variable, NEWS, to proxy for the sign of the news management forecasts convey relative to contemporaneous expectations. *Good news* disclosures include management forecast point estimates above expected earnings, range estimates where the midpoint exceeds expected earnings, and all minimum estimates. Expected earnings equals the median I/B/E/S analyst forecast when available. Otherwise, expected earnings equals the random-walk EPS for annual forecasts and seasonal random-walk for quarterly forecasts. *Bad news* disclosures include point estimates below expected earnings, range estimates where the midpoint is less than expected earnings, and all maximum estimates. NEWS equals 1 for good news forecasts and 0 for bad news forecasts. We discard 32 *neutral news* forecasts (point estimates equaling expected earnings and range estimates where the midpoint equals expected earnings) and all general impression forecasts.¹³

Hypothesis 1 predicts that Canadian firms will be more likely to engage in riskier longer-range forecasting for both bad ($c_2 > 0$) and good ($c_2 + c_3 > 0$) news. Given Skinner's (1994) finding that U.S. firms tend to voluntarily release good news in longer-horizon annual forecasts and bad news in shorter-horizon interim forecasts, we expect $c_1 > 0$, indicating that U.S. good news forecasts are of a longer horizon. Hypothesis 2 predicts that Canadian firms will have less tendency to choose longer forecast horizons for good news relative to U.S. firms ($c_3 < 0$). For the control variables, we expect $c_4 > 0$ (longer forecast horizons for larger firms), $c_5 < 0$ (shorter forecast horizons for regulated firms), and we make no prediction for the coefficient c_6 on high-tech firms.

V. RESULTS

Variable Distributions and Other Descriptive Statistics

Table 2, Panel A describes variable distributions for our sample of 115,751 forecasting periods used to estimate Equation (1) (forecast frequency tests). Mean and median $\Delta\text{EARNINGS}$ are slightly higher than 1 percent and 0.1 percent of share price, respectively. Mean and median $|\Delta\text{EARNINGS}|$ are approximately 8.5 percent and 1.1 percent of price, respectively. Mean earnings change is more positive for Canadian firms (2.15 percent relative to 1.13 percent for U.S. firms), but median earnings change is almost identical (0.16 percent relative to 0.14 percent for U.S. firms). A similar relationship holds for absolute earnings change.¹⁴

Median firm size (per observation) is nearly \$87 million, which is much smaller than in Kasznik and Lev (1995), who report median firm sizes of \$381 and \$436 million for earnings increase and decrease periods, respectively. We do not require I/B/E/S quarterly analyst forecasts for our primary tests, so our sample includes many more small and lightly followed firms. Also, in contrast to Kasznik and Lev (1995), we include only earnings forecasts, not sales forecasts, and we sample from the full year instead of only the fourth quarter where management forecast frequency is the greatest. Not surprisingly then, Panel

¹³ Defining minimum estimates as good news and maximum estimates as bad news is consistent with the stock price reactions to these disclosures documented by Baginski et al. (1994). We also replicated the analysis using a five-level news measure corresponding to the five categories of news in Baginski et al. (1994) with similar results.

¹⁴ Because our hypotheses are based only on the *sign* of the surprise (ΔSIGN), skewness in $\Delta\text{EARNINGS}$ (and its absolute value) only affects the control variables. In untabulated results, we find that log transformation of $|\Delta\text{EARNINGS}|$ does not increase the variable's explanatory power, nor does it change our inferences concerning H1 and H2.

TABLE 2
Variable Distributions for 1993–1996 Sample of 115,751 Potential Forecasting Periods (Panel A) and 1993–1996 Sample of 1,236 Management Earnings Forecasts (Panel B)

Panel A: Variable Distributions for Sample of 115,571 Potential Forecasting Periods and Frequencies and Types of Forecasts Released during those Periods

<i>Continuous Variables</i>	<i>n</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>Lower Quartile</i>	<i>Median</i>	<i>Upper Quartile</i>
ΔEARNINGS	115,751	0.0121	0.2258	−0.0083	0.0014	0.0124
U.S.	106,676	0.0113	0.2237	−0.0082	0.0014	0.0120
Canadian	9,075	0.0215	0.2494	−0.0083	0.0016	0.0182
ΔEARNINGS	115,751	0.0850	0.2096	0.0027	0.0105	0.0462
U.S.	106,676	0.0838	0.2077	0.0027	0.0130	0.0452
Canadian	9,075	0.0991	0.2298	0.0033	0.0130	0.0589
SIZE (millions)	115,751	984.3	4315.3	22.5	86.7	396.1
U.S.	106,676	1001.5	4459.5	22.2	85.1	386.3
Canadian	9,075	782.3	1926.4	29.2	107.3	535.4
<i>Classification Variables</i>	<i>Yes (%)</i>		<i>No (%)</i>		<i>Total (%)</i>	
High-tech firm (HIGHTECH = 1)	20,058 (17.3)		95,693 (82.7)		115,751 (100.0)	
U.S.	18,913 (17.7)		87,763 (82.3)		106,676 (100.0)	
Canadian	1,145 (12.6)		7,930 (87.4)		9,075 (100.0)	
Regulated firm (REGULATE = 1)	15,116 (13.1)		100,635 (86.9)		115,751 (100.0)	
U.S.	14,541 (13.6)		92,135 (86.4)		106,676 (100.0)	
Canadian	575 (6.3)		8,500 (93.7)		9,075 (100.0)	
Good news period (Earnings increase; ΔESIGN = 1)	67,378 (58.2)		48,373 (41.8)		115,751 (100.0)	
U.S.	62,348 (58.4)		44,328 (41.6)		106,676 (100.0)	
Canadian	5,030 (55.4)		4,045 (44.6)		9,075 (100.0)	
<i>Forecast Rates in Potential Forecast Periods</i>	<i>Forecasts</i>		<i>Periods</i>		<i>Rate</i>	
Canadian firms	219		9,075		2.41%	
Good news periods (ΔEARNINGS ≥ 0)	150		5,030		2.98%	
Bad news periods (ΔEARNINGS < 0)	69		4,045		1.71%	
U.S. firms	1,164		106,676		1.09%	
Good news periods	685		52,348		1.10%	
Bad news periods	479		44,328		1.08%	
<i>Management Forecast Type</i>	<i>U.S. Firm</i>		<i>Canadian Firm (%)</i>		<i>Total (%)</i>	
Point	282 (24.2)		101 (46.1)		383 (27.7)	
Range	292 (25.1)		32 (14.6)		324 (23.4)	
Minimum	309 (26.5)		48 (21.9)		357 (25.8)	
Maximum	151 (13.0)		14 (6.4)		165 (12.0)	
General impression	130 (11.2)		24 (11.0)		154 (11.1)	
Total	<u>1,164 (100.0)</u>		<u>219 (100.0)</u>		<u>1,383 (100.0)</u>	

(Continued on next page)

TABLE 2 (Continued)

Panel B: Variable Distributions for 1993–1996 Sample of 1,236 Management Earnings Forecasts for which Forecast News Can Be Calculated (General Impression Forecasts Excluded)

<i>Continuous Variables</i>	<i>n</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>Lower Quartile</i>	<i>Median</i>	<i>Upper Quartile</i>
FHORIZON	1,236	89.2	107.4	6.0	49.0	188.0
U.S.	1,040	70.5	98.1	2.0	30.0	105.5
Canadian	196	188.1	100.3	145.0	215.0	246.0
ALTFHORIZON	1,236	152.5	111.6	63.0	160.0	245.0
U.S.	1,040	142.5	112.9	54.0	139.0	239.5
Canadian	196	205.5	87.6	197.0	228.0	258.0
SIZE (millions)	1,236	1,789.5	4,523.0	74.9	300.7	1,397.8
U.S.	1,040	1,893.3	4,862.4	66.6	280.4	1,374.9
Canadian	196	1,238.5	1,795.9	138.8	541.9	1,526.7
ΔEARNINGS	1,236	0.0343	0.1013	0.0023	0.0080	0.0247
U.S.	1,040	0.0347	0.1079	0.0019	0.0072	0.0221
Canadian	196	0.0324	0.0536	0.0038	0.0113	0.0361
<i>Classification Variables</i>		<i>Yes (%)</i>		<i>No (%)</i>		<i>Total (%)</i>
High-tech firm (HIGHTECH = 1)		211 (17.1)		1,025 (82.9)		1,236 (100.0)
U.S.		193 (18.6)		847 (81.4)		1,040 (100.0)
Canadian		18 (9.2)		178 (90.8)		196 (100.0)
Regulated firm (REGULATE = 1)		45 (3.6)		1,191 (96.4)		1,236 (100.0)
U.S.		39 (3.8)		1,001 (96.2)		1,040 (100.0)
Canadian		6 (3.1)		190 (96.9)		196 (100.0)
Interim forecast		635 (51.4)		601 (48.6)		1,236 (100.0)
U.S.		620 (59.6)		420 (40.4)		1,040 (100.0)
Canadian		15 (7.7)		181 (92.3)		196 (100.0)
Analyst following		910 (73.6)		326 (26.4)		1,236 (100.0)
U.S.		758 (72.9)		282 (27.1)		1,040 (100.0)
Canadian		152 (77.6)		44 (22.4)		196 (100.0)
<i>Forecast News</i>		<i>Total (%)</i>		<i>U.S. (%)</i>		<i>Canadian (%)</i>
Good news (NEWS = 1)		769 (62.2)		650 (62.5)		119 (60.7)
Neutral news (discarded)		32 (2.6)		25 (2.4)		7 (3.6)
Bad news (NEWS = 0)		435 (35.2)		365 (35.1)		70 (35.7)
Total		1,236 (100.0)		1,040 (100.0)		196 (100.0)

FHORIZON = the number of calendar days between the forecast and the end of the period being forecast.

ALTFHORIZON = the number of calendar days between the forecast and the end of the firm's fiscal year;

ΔEARNINGS_{i,t} = [EPS_{i,t} - EPS_{i,t-k}]/PRICE_{i,t-k};

EPS_{i,t} = earnings per share for firm i in period t;

EPS_{i,t-k} = earnings per share for firm i in period t - k;

PRICE_{i,t-k} = security price for firm i in period t - k; and

k = 1 and 4 for annual and quarterly observations, respectively.

ΔESIGN = 1 for ΔEARNINGS ≥ 0 (i.e., good news), 0 for ΔEARNINGS < 0 (i.e., bad news);

|ΔEARNINGS_{i,t}| = absolute value of ΔEARNINGS_{i,t};

(Continued on next page)

TABLE 2 (Continued)

SIZE	= market value of equity at the beginning of the period;
NEWS	= 1 if the forecast is good news, and 0 if the forecast is bad news. <i>Good news</i> disclosures include management forecast point estimates above expected earnings, range estimates where the midpoint exceeds expected earnings, and all minimum estimates. Expected earnings equals the median I/B/E/S analyst forecast when available. Otherwise, expected earnings equals the random-walk EPS for annual forecasts and seasonal random-walk for quarterly forecasts. <i>Bad news</i> disclosures include point estimates below expected earnings, range estimates where the midpoint is less than expected earnings, and all maximum estimates;
HIGHTECH	= 1 if a firm belongs to Pharmaceuticals (Compustat SIC codes 2833–2836), R&D Services (8731–8734), Programming (7371–7379), Computers (3570–3577), or Electronics (3600–3674) industries, and 0 otherwise; and
REGULATE	= 1 if a firm belongs to Telephone (4812–4813), TV (4833), Cable (4841), Communications (4811–4899), Gas (4922–4924), Electricity (4931), Water (4941), or Financial (6021–6023, 6035–6036, 6141, 6311, 6321, 6331) industries, and 0 otherwise.

A shows that forecast frequency is only 1.09 percent for U.S. firms and 2.41 percent for Canadian firms, compared to Kasznik and Lev's (1995) rates of between 9 percent and 21 percent.

U.S. firms report a slightly higher proportion of earnings increases (58.4 percent) during the time period sampled than do Canadian firms (55.4 percent). The U.S. also has a higher percentage of both high-tech and regulated firms. Canadian management earnings forecasts appear in the most precise point form 46.1 percent of the time, nearly twice as often as those of U.S. firms.

Table 2, Panel B presents data for the sample of 1,236 management forecasts possessing all data necessary to estimate Equations (2) through (4) (precision and horizon tests). Panel B's sample equals Panel A's sample of 1,383 management forecasts minus 154 general impression forecasts plus 7 forecasts for which an analyst forecast was available on I/B/E/S when data for constructing a random walk forecast (i.e., prior period EPS) were not available on Compustat. Comparison of Panel B (periods for which the firm issues a forecast) to Panel A (all periods regardless of whether the firm issues a forecast) reveals several expected differences. Firms issuing forecasts are on average much larger; absolute change in earnings during the period ($|\Delta \text{EARNINGS}|$) is much lower; and regulated firms rarely issue forecasts. Panel B also shows that mean (median) forecast horizon (FHORIZON) is 188.1 (215.0) calendar days for Canadian firms but only 70.5 (30.0) days for U.S. firms. U.S. firms are much more apt to issue interim forecasts (59.6 percent of U.S. forecasts) than are Canadian firms (7.6 percent of Canadian forecasts). Even measuring forecast horizon relative to the fiscal year-end, mean (median) ALTFHORIZON is 205.5 (228.0) for Canadian firms but only 142.5 (139.0) days for U.S. firms. Therefore, Canadian firms' annual forecasts occur earlier in the fiscal year than U.S. firms' interim forecasts. In addition, 73.6 percent of the firms are covered by at least one I/B/E/S analyst. The coverage is slightly greater for Canadian firms that issue management forecasts (77.6 percent relative to 72.9 percent for U.S. firms that issue management forecasts). Finally, 62.2 percent of the forecasts are good news. The distributions of NEWS for the U.S. and Canada are fairly similar.

Primary Results: Forecast Frequency

Table 3 presents the Equation (1) logistic regression tests of H1 and H2. Coefficient a_2 is significantly positive ($p = 0.006$), so Canadian firms are more likely to issue forecast

TABLE 3
Management Earnings Forecast Frequency in the U.S. and Canada

$$\text{FORECAST}_{i,t} = a_0 + a_1 \Delta \text{ESIGN}_{i,t} + a_2 \text{CANADA}_{i,t} + a_3 \text{CANADA}_{i,t} \times \Delta \text{ESIGN}_{i,t} \\ + a_4 |\Delta \text{EARNINGS}_{i,t}| + a_5 \text{LSIZE}_{i,t} + a_6 \text{REGULATE}_{i,t} + a_7 \text{HIGHTECH}_{i,t} + \varepsilon_{i,t}$$

<i>Independent Variable (Coefficient)</i>	<i>Expected Sign (Hypothesis)</i>	<i>Coefficient Estimate (p-value)^a</i>
Intercept (a_0)	None predicted	-5.65 (0.001)
ΔESIGN (a_1)	-	-0.14 (0.009)
CANADA (a_2)	+ (H1 for bad news)	0.33 (0.006)
CANADA $\times$ ΔESIGN (a_3)	+ (H2)	0.59 (0.001)
$ \Delta \text{EARNINGS}_{i,t} $ (a_4)	-	-0.52 (0.007)
LSIZE (a_5)	+	0.27 (0.001)
REGULATE (a_6)	-	-1.51 (0.001)
HIGHTECH (a_7)	None predicted	-0.10 (0.149)
Coefficients $a_2 + a_3$	+ (H1 for good news)	0.92 (0.001)
n		115,751
Chi-square		764.8
p-value		(0.001)

^a p-values are one-tailed for signed predictions.

$|\Delta \text{EARNINGS}_{i,t}|$ = absolute value of $\Delta \text{EARNINGS}_{i,t}$, where $\Delta \text{EARNINGS}_{i,t} = [\text{EPS}_{i,t} - \text{EPS}_{i,t-k}]/\text{PRICE}_{i,t-k}$;

$\text{EPS}_{i,t}$ = earnings per share for firm i in period t ;

$\text{EPS}_{i,t-k}$ = earnings per share for firm i in period $t - k$;

$\text{PRICE}_{i,t-k}$ = security price for firm i in period $t - k$; and

$k = 1$ and 4 for annual and quarterly observations, respectively.

$\Delta \text{ESIGN} = 1$ for $\Delta \text{EARNINGS} \geq 0$ (i.e., good news), 0 for $\Delta \text{EARNINGS} < 0$ (i.e., bad news);

$\text{FORECAST} = 1$ if the firm issued a management earnings forecast during the period, and 0 otherwise;

$\text{CANADA} = 1$ if the potential forecasting period relates to a Canadian firm, and 0 if it relates to a U.S. firm;

LSIZE = natural log of the market value of equity at the beginning of the period;

$\text{HIGHTECH} = 1$ if a firm belongs to Pharmaceuticals (Compustat SIC codes 2833–2836), R&D Services (8731–8734), Programming (7371–7379), Computers (3570–3577), or Electronics (3600–3674) industries, and 0 otherwise; and

$\text{REGULATE} = 1$ when a firm belongs to Telephone (4812–4813), TV (4833), Cable (4841), Communications (4811–4899), Gas (4922–4924), Electricity (4931), Water (4941), or Financial (6021–6023, 6035–6036, 6141, 6311, 6321, 6331) industries, and 0 otherwise.

disclosures during bad news periods relative to U.S. firms. The sum $a_2 + a_3$ is also significantly positive ($p < 0.001$), revealing that Canadian firms are more likely to issue forecast disclosures during good news periods relative to U.S. firms. These results support H1's prediction that lower legal liability in Canada leads to more forecast disclosure for *both* good and bad news periods.¹⁵ These results are consistent with the higher univariate forecast rates for Canadian firms (2.41 percent) relative to U.S. firms (1.09 percent) reported in Table 2, Panel A.

With respect to H2, the results are also consistent with our expectations. The tendency to skew disclosure toward bad news periods relative to good news periods differs between

¹⁵ Untabulated analysis reveals that forecasting frequencies for Canadian cross-listed firms (2.02 percent) are smaller than for Canadian locally listed firms (3.01 percent), but greater than for U.S. firms (1.09 percent). The ordering also holds for bad news and good news considered separately.

the U.S. and Canada. U.S. firm behavior is as expected: coefficient a_1 is significantly negative ($p = 0.009$), indicating more forecast disclosure in bad news periods relative to good news periods. Coefficient a_3 , which measures the difference between Canadian and U.S. sign-related behavior, is significantly positive ($p < 0.001$), indicating that Canadian firms are less apt to skew disclosure toward bad news periods than are U.S. firms.¹⁶ In fact, Table 3 reveals that Canadian firms skew disclosure toward good news. The sum $a_1 + a_3$ measures the extent to which Canadian forecast frequency is related to the sign of earnings changes (i.e., the sum $a_1 + a_3$ is the Canadian equivalent to coefficient a_1 for U.S. firms). This sum is positive and significant ($p < 0.001$; results not tabulated), indicating that Canadian forecasts occur more often in good news periods. The univariate forecast rates reported in Table 2, Panel A are consistent with this result.¹⁷

These differences between Canadian and U.S. forecasting behavior arise even after controlling for the absolute value of the earnings surprise (a_4 is significantly negative as expected, $p = 0.007$), firm size (a_5 is significantly positive as expected, $p < 0.001$), regulated industry membership (a_6 is significantly negative as expected, $p < 0.001$), and high-tech industry membership (a_7 is insignificantly different from 0).¹⁸

Primary Results: Forecast Precision

Table 4 presents results of estimating Equation (2). As H1 predicts, the significantly positive b_2 ($p < 0.001$) indicates that Canadian firms issue more precise forecasts in bad news periods than do U.S. firms, and the significantly positive sum $b_2 + b_3$ ($p < 0.001$) indicates that Canadian firms also issue more precise forecasts in good news periods than do U.S. firms. Coefficient b_1 on ΔESIGN is significantly positive ($p = 0.028$), indicating

¹⁶ Table 3 results could also occur if DJNRS selectively judges U.S. good news as insufficiently newsworthy to report. Such a selective bias would make it appear as if (1) more Canadian forecasts existed (because some U.S. forecasts were not reported), and (2) sign-related behavior existed for U.S. firms (because editors biased reporting toward bad news), but not for Canadian firms (because the editors' bias existed for only U.S. firms). Additional analysis suggests that this alternative interpretation is unlikely to explain our results.

PR Newswire carries all subscribing company disclosures, and thus is free of editorial bias. We searched PR Newswire for one of our sample years (1993) and found 89 management forecasts for subscribing companies. Then we used DJNRS to determine whether we would pick up these forecasts. Of the 89 forecasts from PR Newswire, we located 83 (93 percent). Of the six that we could not locate, four were bad news (i.e., earnings lower than previous expectations) and two were good news (i.e., record earnings). This analysis suggests that Dow Jones picks up substantially all of the information contained in PR Newswire. To the extent that Dow Jones is missing any information, it does not appear that it is a result of an editorial bias in favor of reporting bad news. These findings suggest that our evidence is not attributable to Dow Jones editorial policies that favor the dissemination of bad news over good news for U.S. firms.

¹⁷ In an untabulated supplemental analysis, we split the full sample into annual and interim periods, which permits a closer comparison to Skinner's (1994) U.S. results. Coefficient a_1 is significantly positive in the annual sample ($p = 0.028$), but significantly negative in the interim sample ($p < 0.001$). This result corroborates (in our larger and more diverse sample) Skinner's (1994) finding that U.S. managers are more likely to issue interim forecasts when earnings are declining, while they are more likely to issue annual forecasts when earnings are increasing. Canadian firms issue more forecasts in both good and bad news periods relative to U.S. firms only in the annual subsample ($p < 0.001$ for annual regression coefficients a_2 and $a_2 + a_3$). Canadian firms are less likely to skew forecast incidence toward earnings decrease periods (relative to U.S. firms) only in the interim sample ($a_3 > 0$; $p = 0.048$). These results suggest that Canadian firms, operating in a less litigious environment, are issuing more forecasts, but are not using interim forecasts to convey bad news.

¹⁸ We examined other potential control variables for which data were available for subsets of our sample. Substantial differences in earnings capitalization rates affect both the costs and benefits of a publicly released forecast by affecting the price reaction to a given dollar of unexpected earnings. Consequently, we examined whether it was necessary to control for differences in earnings capitalization rates and growth using a subset of firms for which data were available (results not tabulated). Specifically, including as an additional variable ranked, leading price/earnings ratios (for firms with positive leading earnings) did not affect our conclusions. Also, proprietary information could adversely affect management's willingness to release an earnings forecast publicly. In addition to our control for high-tech firms, we added another proxy for proprietary information, expected growth (measured as market to book value of common equity), with no change in our conclusions.

TABLE 4
Determinants of Management Forecast Precision in the U.S. and Canada

$$\text{Model: } \text{PRECISE}_i = b_0 + b_1 \Delta \text{ESIGN}_i + b_2 \text{CANADA}_i + b_3 \text{CANADA}_i \times \Delta \text{ESIGN}_i + b_4 \text{LSIZE}_i \\ + b_5 \text{REGULATE}_i + b_6 \text{HIGHTECH}_i + b_7 \text{FHORIZON}_i + v_i$$

<i>Independent Variable (Coefficient)</i>	<i>Expected Sign (Hypothesis)</i>	<i>Coefficient Estimate (p-value^a)</i>
ΔESIGN (b_1)	+	0.211 (0.028)
CANADA (b_2)	+ (H1 for bad news)	1.060 (0.001)
CANADA $\times$ ΔESIGN (b_3)	– (H2)	–0.141 (0.301)
LSIZE (b_4)	–	–0.007 (0.393)
REGULATE (b_5)	–	–0.247 (0.155)
HIGHTECH (b_6)	None predicted	–0.178 (0.090)
FHORIZON (b_7)	–	–0.003 (0.001)
Coefficients $b_2 + b_3$	+ (H1 for good news)	0.919 (0.001)
n		1,383
Model Chi-square probability		67.13 (0.001)

^a p-values are one-tailed for signed predictions.

PRECISE = 3 (2, 1, 0) for point (range, open-interval, general impression) forecasts;

$\Delta \text{EARNINGS}_{i,t} = [\text{EPS}_{i,t} - \text{EPS}_{i,t-k}] / \text{PRICE}_{i,t-k}$;

$\text{EPS}_{i,t}$ = earnings per share for firm i in period t ;

$\text{EPS}_{i,t-k}$ = earnings per share for firm i in period $t - k$;

$\text{PRICE}_{i,t-k}$ = security price for firm i in period $t - k$; and

$k = 1$ and 4 for annual and quarterly observations, respectively.

$\Delta \text{ESIGN} = 1$ for $\Delta \text{EARNINGS} \geq 0$ (i.e., good news), 0 for $\Delta \text{EARNINGS} < 0$ (i.e., bad news);

CANADA = 1 if a Canadian firm issued the forecast, and 0 if a U.S. firm issued the forecast;

LSIZE = natural log of the market value of equity at the beginning of the period;

HIGHTECH = 1 if a firm belongs to Pharmaceuticals (Compustat SIC codes 2833–2836), R&D Services (8731–8734), Programming (7371–7379), Computers (3570–3577), or Electronics (3600–3674) industries, and 0 otherwise;

REGULATE = 1 if a firm belongs to Telephone (4812–4813), TV (4833), Cable (4841), Communications (4811–4899), Gas (4922–4924), Electricity (4931), Water (4941), or Financial (6021–6023, 6035–6036, 6141, 6311, 6321, 6331) industries, and 0 otherwise; and

FHORIZON = the number of calendar days between the forecast and the end of the period being forecast.

that U.S. firms on average issue less precise forecasts when earnings are declining. However, the data do not support H2's prediction that the sign of earnings change has less effect on forecast precision in Canada (b_3 is insignificant, $p = 0.301$). These results arise after controlling for the expected negative effects of long forecast horizons on forecast precision (negative b_7 ; $p < 0.001$), high-tech firms that issue less precise forecasts (negative b_6 ; $p = 0.090$), and firm size and regulated firms (insignificant b_4 and b_5 , respectively). In

summary, the results reported in Table 4 indicate that Canadian firms issue more precise forecasts than do U.S. firms.

Primary Results: Forecast Horizon

Table 5 presents OLS regression estimates of Equations (3) and (4), designated Model 1 and Model 2, respectively. Hypothesis 1 predicts that Canadian firms issue longer-horizon forecasts. Consistent with H1, the significantly positive c_2 ($p < 0.001$ in both models), indicates that Canadian firms issue longer-horizon bad news forecasts than do U.S. firms, and the significantly positive sum $c_2 + c_3$ ($p < 0.001$ in both models), indicates that

TABLE 5
Determinants of Management Forecast Horizons in the U.S. and Canada

$$\text{Model 1: } FHORIZON_i = c_0 + c_1 NEWS_i + c_2 CANADA_i + c_3 CANADA_i \times NEWS_i \\ + c_4 LSIZE_i + c_5 REGULATE_i + c_6 HIGHTECH_i + \mu_i$$

$$\text{Model 2: } ALTFHORIZON_i = c_0 + c_1 NEWS_i + c_2 CANADA_i + c_3 CANADA_i \times NEWS_i \\ + c_4 LSIZE_i + c_5 REGULATE_i + c_6 HIGHTECH_i + \eta_i$$

Independent Variable (Coefficient)	Expected Sign (Hypothesis)	Estimated Coefficients (p-values ^a)	
		Model 1	Model 2
Intercept (c_0)	None predicted	31.69 (0.001)	110.76 (0.001)
NEWS (c_1)	+	30.61 (0.001)	11.67 (0.050)
CANADA (c_2)	+ (H1 for bad news)	141.82 (0.001)	80.78 (0.001)
CANADA $\times$ NEWS (c_3)	– (H2)	–35.18 (0.013)	–33.34 (0.031)
LSIZE (c_4)	+	3.78 (0.004)	4.86 (0.002)
REGULATE (c_5)	–	12.35 (0.800)	20.43 (0.891)
HIGHTECH (c_6)	None predicted	–15.47 (0.038)	–18.91 (0.012)
Coefficients $c_2 + c_3$	+ (H1 for good news)	106.64 (0.001)	47.44 (0.001)
n		1,204	1,204
Adjusted R ²		0.1931	0.0556
Model F		48.97	12.80

^a p-values are one-tailed for signed predictions.

FHORIZON = the number of calendar days between the forecast and the end of the period being forecast;

ALTFHORIZON = the number of calendar days between the forecast and the end of the firm's fiscal year;

NEWS = 1 if the forecast is good news, and 0 if the forecast is bad news. *Good news* disclosures include management forecast point estimates above expected earnings, range estimates where the midpoint exceeds expected earnings, and all minimum estimates. Expected earnings equals the median I/B/E/S analyst forecast when available. Otherwise, expected earnings equals the random-walk EPS for annual forecasts and seasonal random-walk for quarterly forecasts. *Bad news* disclosures include point estimates below expected earnings, range estimates where the midpoint is less than expected earnings, and all maximum estimates;

CANADA = 1 if a Canadian firm issued the forecast, and 0 if a U.S. firm issued the forecast;

LSIZE = natural log of the market value of equity at the beginning of the period;

HIGHTECH = 1 if a firm belongs to Pharmaceuticals (Compustat SIC codes 2833–2836), R&D Services (8731–8734), Programming (7371–7379), Computers (3570–3577), or Electronics (3600–3674) industries, and 0 otherwise; and

REGULATE = 1 if a firm belongs to Telephone (4812–4813), TV (4833), Cable (4841), Communications (4811–4899), Gas (4922–4924), Electricity (4931), Water (4941), or Financial (6021–6023, 6035–6036, 6141, 6311, 6321, 6331) industries, and 0 otherwise.

Canadian firms also issue longer-horizon good news forecasts than do U.S. firms.¹⁹ U.S. firms forecast good news over longer horizons relative to bad news (significantly positive c_1 , $p < 0.05$). Further, the data also support the H2 prediction that the sign of forecast news has less effect on Canadian firms' forecast horizon. As expected, Canadian firms are less apt to choose longer forecast horizons for good news relative to bad news (significantly negative c_3 , $p = 0.031$).

VI. SENSITIVITY CHECKS

Effects of the Private Securities Litigation Reform Act on U.S. Forecasting Behavior

On December 22, 1995, Congress voted to override a presidential veto, thereby establishing the Private Securities Litigation Reform Act of 1995. One intent of the Act was to encourage firms to disclose forward-looking information by reducing legal penalties. If management forecasting behavior changed after the Act, then including post-Act U.S. forecasts in our sample biases the results in favor of the null hypothesis of no differences between U.S. and Canadian forecasting behavior. We reject the null hypothesis in each of our tests except for H2 in Table 4, where we find no difference between Canada and the U.S. in the association of the sign of the earnings change and forecast precision. Accordingly, we ran a model analogous to Equation (2) on U.S. firms only, replacing the Canadian regime indicator variable (CANADA) with a 1,0 indicator variable for whether the forecast was issued after the Act in 1996. We find that, in 1996, U.S. forecasting firms still issue less precise forecasts in periods of earnings decreases ($p = 0.022$; results not tabulated). However, running Equation (2) without 1996 U.S. forecasts does not change our inference of no significant difference between Canada and the U.S. in the association of the sign of the earnings change and forecast precision.

Information Asymmetry

The benefits of voluntary management forecast disclosure increase when private information exists (King et al. 1990). Fowler et al. (1979) and Jorion and Schwartz (1986) document thin-trading for many TSE firms, a symptom of information asymmetry.²⁰ If Canadian managers wish to reduce such asymmetry to decrease the firm's cost of capital (King et al. 1990), then they will be more likely to issue management earnings forecasts, and the forecasts will on average be longer-horizon. Although Canadian managers may also wish to issue more precise forecasts when faced with greater information asymmetry, their forecasts will likely be less precise because information asymmetry is strongly linked to uncertainty, and Baginski and Hassell (1997) document that management forecast precision is inversely related to uncertainty.

We replicated our entire study for the subset of firm-periods for which more than one analyst follows the firm at the beginning of the period. We use the standard deviation of these analyst forecasts as a proxy for predisclosure information asymmetry (e.g., Atiase and Bamber 1994). Specifically, we divide the standard deviation by beginning of the period price, rank the deflated variable, and include it as an additional control variable. Greater

¹⁹ Cross-listed Canadian firms' forecast horizons fall in-between the forecast horizons of Canadian locally listed and U.S. firms. Horizons in calendar days for Canadian locally listed, Canadian cross-listed, and U.S. forecasts are 194, 182, and 70 days, respectively. For the alternative definition of forecast horizon relative to the fiscal year-end, the lengths are 208, 202, and 142 days, respectively.

²⁰ As firm size and analyst following increase, TSE thin-trading becomes a lesser issue. Fowler et al. (1979) note that of the firms in the TSE 300 index, approximately only 6 percent trade infrequently (trades on fewer than 80 percent of sampled days).

information asymmetry leads to greater forecasting benefits (i.e., removal of information asymmetry). Thus, our proxy should be associated with more management forecasts, longer forecast horizons, and due to its link to uncertainty, with less precise forecasts. The results (not tabulated) are consistent with these expectations, but our conclusions concerning H1 and H2 remain unchanged.²¹

VII. SUMMARY AND CONCLUSIONS

Policymakers regularly call for U.S. corporations to increase disclosure of forward-looking information (e.g., AICPA 1994; Breeden 1995). In this paper, we show that the frequency and characteristics of voluntary earnings forecast disclosures differ across two countries with different legal regimes. Canadian managers, faced with a less litigious environment than U.S. managers, disclose more earnings forecasts (in periods of both good and bad news), and their forecasts are more precise and of longer range. In addition, using a larger and more representative sample of firms, we confirm the findings in prior research that U.S. forecasts occur more often, are less precise, and are of a shorter horizon when earnings news is bad. In contrast, we document that Canadian forecasts are less likely to exhibit these bad-news-related tendencies. In fact, Canadian firms issue more forecasts in earnings increase relative to earnings decrease periods.

Our evidence that lower legal liability is consistent with more voluntary management forecast disclosure is consistent with evidence in Frost (2001) and Johnson et al. (2001). Our finding that Canadian firms issue proportionally more good news forecasts in the lower legal liability regime is consistent with Skinner's (1994) and Kasznik and Lev's (1995) conclusions that legal exposure leads to asymmetric incentives favoring proportionally more disclosure of bad news. However, the finding is inconsistent with Johnson et al. (2001), who fail to find a proportional decrease in bad news forecasts (or proportional increase in good news forecasts) after passage of the Private Securities Litigation Reform Act that was intended to reduce legal liability. This pattern of evidence suggests that either (1) the difference in legal exposure between the U.S. and Canada is greater than the change in legal exposure in the U.S. brought about by the Private Securities Litigation Reform Act, or that (2) post-Act behavior of firms in high-tech industries differs substantially from that of firms in other industries.

Although we are unaware of any documented biases in our data sources that could confound our analysis, sampling of periods available for disclosure and actual forecast disclosures is limited by the completeness of sources such as Compustat and DJNRS. Although our sample is from the relatively recent 1993–1996 period, our results may not be generalizable to the present. U.S. voluntary disclosure policy is constantly evolving, and managers may now have more evidence on whether the Private Securities Litigation Reform Act of 1995 actually reduces disclosure-related legal exposure.

Our evidence suggests that legal regime is associated with management forecasting behavior. Future research could assess whether lower legal liability undermines the credibility of Canadian management forecasts. To date, research on the relation between legal liability and forecast credibility focuses on the passage of the Private Securities Litigation Reform Act of 1995, and its results are mixed. Johnson et al. (2001) find no change in management forecast accuracy and bias post-Act; however, Ali and Kallapur's (2001) results suggest that shareholders reacted negatively to the Act's restrictions on their ability

²¹ Our conclusions are also unaffected when we control for the number of analysts who follow the firm (results not tabulated).

to bring suit. Substantial difference in legal regimes across countries might provide a powerful treatment effect for detecting differences in forecast credibility and shareholders' perceptions of how their wealth is affected by protections under different legal regimes.

REFERENCES

- Ajinkya, B., and M. Gift. 1984. Corporate managers' earnings forecasts and symmetrical adjustments of market expectations. *Journal of Accounting Research* 22 (Autumn): 425–444.
- Ali, A., and L. Hwang. 2000. Country-specific factors related to financial reporting and the value relevance of accounting data. *Journal of Accounting Research* 38 (Spring): 1–21.
- , and S. Kallapur. 2001. Securities price consequences of the Private Securities Litigation Reform Act of 1995 and related events. *The Accounting Review* 76 (July): 431–460.
- American Institute of Certified Public Accountants (AICPA), Special Committee on Financial Reporting. 1994. *Improving Business Reporting—A Customer Focus*. New York, NY: AICPA.
- Atiase, R., and L. S. Bamber. 1994. Trading volume reactions to annual accounting earnings announcements: The incremental role of predisclosure information asymmetry. *Journal of Accounting and Economics* 17 (May): 309–329.
- Baginski, S., J. Hassell, and G. Waymire. 1994. Some evidence on the news content of preliminary earnings estimates. *The Accounting Review* 69 (January): 265–273.
- , and ———. 1997. Determinants of management forecast precision. *The Accounting Review* (April): 303–312.
- Bamber, L. S., and Y. Cheon. 1998. Discretionary management earnings forecast disclosures: Antecedents and outcomes associated with forecast venue and forecast specificity choices. *Journal of Accounting Research* 36 (Autumn): 167–190.
- Bandyopadhyay, S., J. D. Hanna, and G. Richardson. 1994. Capital market effects of U.S.-Canada GAAP differences. *Journal of Accounting Research* 32 (Autumn): 262–277.
- Barth, M., and G. Clinch. 1996. International accounting differences and their relation to share prices: Evidence of U.K., Australian, and Canadian firms. *Contemporary Accounting Research* (Spring): 135–170.
- Breeden, R. 1995. Testimony before the Securities Subcommittee of the Senate Committee on Banking, Housing, and Urban Affairs. April 6, 1995. Washington, D.C.: Government Printing Office.
- Brown, L., G. Richardson, and C. Trzcinka. 1991. Strong-form efficiency on the Toronto Stock Exchange: An examination of analyst price forecasts. *Contemporary Accounting Research* 7: 323–346.
- Clarkson, P., A. Dontoh, G. Richardson, and S. Sefcik. 1992. The voluntary inclusion of earnings forecasts in IPO prospectuses. *Contemporary Accounting Research* (Spring): 601–626.
- , and D. Simunic. 1994. The association between audit quality, retained ownership, and firm-specific risk in U.S. vs. Canadian IPO markets. *Journal of Accounting and Economics* 17 (January): 207–228.
- Core, J. 1997. On the corporate demand for directors' and officers' insurance. *Journal of Risk and Insurance* 64 (March): 63–87.
- Cox, C. 1985. Further evidence on the representativeness of management earnings forecasts. *The Accounting Review* 60 (October): 692–701.
- Francis, J., D. Philbrick, and K. Schipper. 1994. Shareholder litigation and corporate disclosures. *Journal of Accounting Research* 32 (Autumn): 137–164.
- Fowler, D., C. H. Rorke, and V. Jog. 1979. Heteroscedasticity, R^2 and thin trading on the Toronto Stock Exchange. *Journal of Finance* 34: 1201–1210.
- Freund, W., M. Larrain, and M. Pagano. 1997. Market efficiency before and after the introduction of electronic trading at the Toronto Stock Exchange. *Review of Financial Economics* 6: 29–56.
- Frost, C., and G. Pownall. 1994. Accounting disclosure practices in the United States and the United Kingdom. *Journal of Accounting Research* 32 (Spring): 75–102.
- , and W. Kinney, Jr. 1996. Disclosure choices of foreign registrants in the United States. *Journal of Accounting Research* 34 (Spring): 67–84.

- . 2001. Characteristics and information value of corporate disclosures of forward-looking information in global equity markets. Working paper, Dartmouth College.
- Grossman, B. 1996. Class rules. *CA Magazine* 129 (March): 31–33, 43.
- Grundfest, J., and M. Perino. 1997. Security litigation reform: The first years' experience. Working paper, Stanford University.
- Guenther, D., and A. Rosman. 1994. Differences between COMPUSTAT and CRSP SIC codes and related effects on research. *Journal of Accounting and Economics* 18 (July): 115–128.
- Johnson, M., R. Kasznik, and K. Nelson. 2001. The impact of securities litigation reform on the disclosure of forward-looking information by high technology firms. *Journal of Accounting Research* 39 (September): 297–327.
- Jorion, P., and E. Schwartz. 1986. Integration vs. segmentation in the Canadian stock market. *Journal of Finance* 41 (July): 603–614.
- Kasznik, R., and B. Lev. 1995. To warn or not to warn: Management disclosures in the face of an earnings surprise. *The Accounting Review* 70 (January): 113–134.
- King, R., G. Pownall, and G. Waymire. 1990. Expectations adjustment via timely management forecasts: Review, synthesis, and suggestions for future research. *Journal of Accounting Literature* 9: 113–144.
- Levine, D., and A. Pritchard. 1998. The Securities Litigation Uniform Standards Act of 1998: The sun sets on California's Blue Sky Laws. *Business Lawyer* 54 (November): 1–54.
- L'Her, J., and J. Suret. 1991. The reaction of Canadian securities to revisions of earnings forecasts. *Contemporary Accounting Research* (Spring): 378–406.
- McConomy, B. 1998. Bias and accuracy of management earnings forecasts: An evaluation of the impact of auditing. *Contemporary Accounting Research* 15 (Summer): 167–195.
- McNichols, M. 1989. Evidence of informational asymmetries from management earnings forecasts and stock returns. *The Accounting Review* 64 (January): 1–27.
- Paskell-Mede, M. 1994. What liability crisis? *CA Magazine* 127 (May): 42–43.
- . 1999. Cross-border lawsuits. *CA Magazine* 132 (April): 37–38.
- Patell, J. 1976. Corporate forecasts of earnings per share and stock price behavior: Empirical tests. *Journal of Accounting Research* (Autumn): 246–276.
- Skinner, D. 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research* 32 (Spring): 38–60.
- . 1997. Earnings disclosure and stockholder lawsuits. *Journal of Accounting and Economics* 23 (November): 249–282.
- Toronto Stock Exchange (TSE). 2000. *The Toronto Stock Exchange's Policy Statement on Timely Disclosure*. Toronto, Ontario.
- Waymire, G. 1984. Additional evidence on the information content of management earnings forecasts. *Journal of Accounting Research* 22 (Autumn): 703–718.
- . 1985. Earnings volatility and voluntary management forecast disclosure. *Journal of Accounting Research* 23 (Spring): 268–295.

Predecisional Distortion of Evidence as a Consequence of Real-Time Audit Review

T. Jeffrey Wilks
Brigham Young University

ABSTRACT: With the current shift toward real-time audit review, subordinates become aware of supervisors' views earlier in the audit process. I use an experiment to examine whether earlier knowledge of supervisors' views increases subordinates' tendencies to agree with those views because subordinates predecisionally distort evidence. In a going-concern task, I find that auditors who learn the partner's view *before* evaluating evidence (1) evaluate individual evidence items as more consistent with the partner's view, and (2) make going-concern judgments that are more consistent with the partner's view, than do auditors who learn the same partner's view *after* evaluating evidence. In a second experiment, I examine whether auditors anticipate the distortion's effect on subordinates' judgments. I find that auditors expect subordinates to make judgments that agree with supervisors' views, but auditors do not expect subordinates to agree even more with those views when subordinates learn those views earlier in the audit process.

Keywords: *predecisional distortion of information; motivated reasoning; accountability; real-time audit review.*

Data Availability: *Contact the author.*

This paper is based on my dissertation completed at Cornell University. I express special thanks to my committee members, Mark Nelson, Jay Russo, Naomi Altman, and especially my chair, Bob Libby, for their invaluable guidance and encouragement. I also thank Jeffrey Hales for his help in collecting data and Ron Guymon for his research assistance. This paper has benefited from the written comments of Steve Asare, Brian Ballou, Mike Bamber, Bob Bowen, Cliff Brown, Tom Dyckman, Steve Glover, Jeffrey Hales, Pat Hopkins, Jane Kennedy, Susan Krische, Marlys Lipe, Laureen Maines, Linda Meyers, Mark Peecher, Fred Phillips, Jamie Pratt, Doug Prawitt, Nathan Stuart, Robin Tarpley, Jim Wahlen, Mark Zimbelman, two anonymous reviewers, and workshop participants at Brigham Young University, Cornell University, the University of Florida, the University of Georgia, the University of Illinois, Indiana University, Oklahoma University, the University of Washington, the 2000 American Accounting Association Annual Meeting, and the 14th Symposium on Auditing Research at the University of Illinois. I am grateful to the Deloitte & Touche Foundation and the American Accounting Association for their generous financial support, and to the participating accounting firm for their cooperation and assistance. This research was supported in part by National Science Foundation Grant SBR-9617873.

Submitted June 2000
Accepted June 2001

I. INTRODUCTION

The audit-review process is shifting from a sequential process to a more real-time process, wherein supervisors give subordinates earlier and more frequent guidance to expedite field work, reduce back-end review, increase interaction among team members, and accelerate issue identification and resolution (Rich et al. 1997a). While the attention-directing effects of guidance can help achieve these goals (e.g., Libby and Trotman 1993; Tan 1995), the time when supervisors provide guidance relative to the time subordinates acquire and evaluate audit evidence can adversely affect subordinates' judgments. For example, prior research shows that subordinates tend to make judgments that deliberately or strategically agree with supervisors' views (Tetlock 1992; Lerner and Tetlock 1999). I expect this tendency to agree to be even more pronounced when supervisors give guidance before (instead of after) subordinates evaluate evidence, because subordinates unconsciously interpret evidence in a manner that is overly consistent with supervisors' views—a phenomenon the psychology literature labels “predecisional distortion of information” (Russo et al. 1996).

In this study, I examine whether disclosure of a supervisor's views earlier in the review process increases subordinates' tendency to agree with those views because subordinates predecisionally distort evidence. Although prior studies demonstrate that supervisors' views (e.g., Peecher 1996) and clients' views (e.g., Cuccia et al. 1995; Ponemon 1995; Buchman et al. 1996; Hackenbrack and Nelson 1996) influence subordinates' judgments, none of these studies examines the incremental effect of predecisional distortion. Instead, they examine how these views influence judgments through subordinates' hypothesis generation (e.g., Libby 1985; Heiman 1990; Church and Schneider 1993), evidence search (e.g., Brown et al. 1999; Cloyd and Spilker 1999), and evidence weighting (e.g., Church 1991; Johnson 1993; McMillan and White 1993; Boiney et al. 1997). While the motivated-reasoning literature highlights all of these (including distortion) as potential cognitive processes through which incentives can influence judgments (Pyszczynski and Greenberg 1987; Kunda 1990), accounting research has not considered the unconscious process of distortion.

Understanding predecisional distortion is important because corrective interventions that address shortcomings in hypothesis generation, evidence search, or evidence weighting are unlikely to mitigate shortcomings due to predecisional distortion (e.g., Libby 1981, 4; Libby and Luft 1993). For example, if supervisors' views inappropriately influence subordinates' judgments because subordinates assign more weight to evidence items that are consistent with those views and less weight to items that are inconsistent with those views, then decision aids that affect the weighting of evidence (e.g., Libby and Libby 1989; Bonner et al. 1996) are likely to prevent unwarranted agreements. However, if subordinates unconsciously evaluate the diagnosticity of an evidence item as overly consistent with supervisors' views, then decision aids that affect the weighting of evidence items are unlikely to prevent unwarranted agreements because the diagnostic values themselves are distorted. In this case, corrective interventions must address how auditors evaluate evidence items, in addition to how they weight those items.

Examining predecisional distortion is also important because prior distortion research relies on students making product-choice decisions (Russo et al. 1996; Russo et al. 1998) or on professionals making judgments on issues that do not require their expertise (Russo et al. 2000). It is an open empirical question whether experienced professionals such as auditors distort evidence in judgments that require their expertise. In fact, Russo et al. (2000) provide evidence that audit seniors making even *nonexpert* judgments distort information significantly less than do pharmaceutical salespersons. Thus, it is unclear whether experts making expert decisions predecisionally distort information. Moreover, because of

the experimental designs used in these prior studies, it is also unclear whether predecisionally distorted information actually influences final judgments.

In Experiment 1, audit managers from a Big 5 accounting firm evaluate 20 evidence items related to a client's financial viability. Either before or after evaluating this evidence, managers read that the engagement partner thought prior auditors were either unduly optimistic or unduly pessimistic when making going-concern judgments. All managers then make going-concern judgments. Consistent with prior research, I find that managers make going-concern judgments that agree with the partner's statement (e.g., Peecher 1996). More importantly, I find that managers who read the partner's statement *before* evaluating evidence interpret individual evidence items in a direction consistent with the partner's statement. As a result, these managers later make going-concern judgments that are more consistent with the partner's statement than do managers who read the same statement *after* evaluating evidence. These results suggest that even experienced auditors making expert judgments predecisionally distort evidence, and this distorted evidence does indeed influence their final judgments.

The impact of predecisional distortion on overall audit conclusions depends on whether supervisors anticipate such behavior from subordinates. Although supervisors' knowledge of persuasion tactics is likely to help them anticipate a subordinate's deliberate agreement with their views (Rich et al. 1997b), supervisors are unlikely to have learned to anticipate distortion effects because distortion itself is largely unconscious (Russo et al. 2000). In a second experiment, I examine whether auditors anticipate both the deliberate and the unconscious effects a supervisor's view can have on subordinates' judgments. Specifically, I ask auditors from a Big 5 accounting firm to anticipate the going-concern judgment a new manager would make given the scenario presented in Experiment 1. In a between-subjects design, I manipulate the partner's view and the time when the manager learned the partner's view—either before or after the manager evaluated evidence. I find that auditors expect the manager to make a going-concern judgment that agrees with the partner's view. However, auditors do not expect the manager to agree even more with the partner's view when that view is known earlier in the audit process. A follow-up question suggests that simply asking auditors to focus on the time when the manager learns the partner's view triggers the anticipation of this increased tendency for subordinates to agree. This suggests that training is likely to help auditors anticipate the predecisional-distortion effects supervisors' views can have on subordinates' judgments.

II. LITERATURE ANALYSIS AND HYPOTHESES

Accountability and Motivated Reasoning

Prior accountability research demonstrates that subordinates tend to make judgments that agree with supervisors' or clients' views (Tetlock 1992; Lerner and Tetlock 1999). Although prior research has documented such behavior in a number of accounting settings, including analytical review (Peecher 1996), interpretation of tax standards (Cuccia et al. 1995), financial reporting (Hackenbrack and Nelson 1996), litigation disclosure (Buchman et al. 1996), and inventory valuation (Ponemon 1995),¹ few of these studies directly examine the cognitive processes through which supervisors' views influence subordinates' judgments. In fact, although prior research thoroughly examines how accountability to audiences with

¹ A number of other studies presume that auditors make judgments that agree with their clients' preferences. These studies hold the client preference constant and examine whether factors such as the ambiguity of case precedents (Nelson and Kinney 1997; Salterio and Koonce 1997), engagement pressure, or appropriateness considerations (Kadous et al. 2001) mitigate the tendency for auditors to agree with client preferences.

unknown views influences cognition (e.g., Kennedy 1993, 1995; see Lerner and Tetlock [1999] for review), studies that examine accountability to audiences with *known* views generally conclude that individuals:

adopt positions likely to gain the favor of those to whom they are accountable, thereby allowing them to avoid the unnecessary cognitive work of analyzing the pros and cons of alternative courses of action, interpreting complex patterns of information, and making difficult trade-offs. (Lerner and Tetlock 1999, 256)

Such a conclusion suggests that subordinates who know their supervisors' views waste little time actually analyzing information, but instead deliberately agree with those views. In contrast, the motivated-reasoning literature suggests that a desire to please or persuade (Rich et al. 1997b) a supervisor also influences subordinates' judgments through cognitive processes such as hypothesis generation, selection of inference rules, evidence search, evidence evaluation, and evidence weighting (Pyszczynski and Greenberg 1987; Kunda 1990). As indicated earlier, prior accounting research has focused on hypothesis generation, evidence search, and evidence weighting, but not on how supervisors' views influence subordinates' evaluation of evidence.

Predecisional Distortion of Information

Research on predecisional distortion views decision making as a sequential discrimination process, wherein individuals (1) acquire a unit of information, typically an attribute or feature of the alternatives being considered, (2) evaluate its diagnosticity or discriminative value, and (3) combine the resulting evaluation with all prior knowledge to create an updated cumulative preference or judgment (Russo et al. 2000, 14). Predecisional distortion of information occurs when individuals unknowingly evaluate the diagnosticity or discriminative value of newly acquired information to favor a previously held belief, desired outcome, or developing preference. This distortion of information occurs before the individual weights or combines the information with prior knowledge to arrive at an updated judgment. For example, Russo et al. (1996) manipulate students' prior preferences for one of two restaurants or one of two inns and find that students rate new product information as more favorable to the preferred alternative. Later, Russo et al. (1998) demonstrate that in the absence of a manipulated prior preference, even a developing preference can trigger distorted ratings of product information.

The predecisional-distortion literature predicts that subordinates who know their supervisors' views are likely to evaluate evidence as overly consistent with those views.² However, most of this research relies on students making product-choice decisions (Russo et al. 1996; Russo et al. 1998) or on professionals making judgments on issues that do not require their expertise (Russo et al. 2000). In reviewing the research on professional auditor judgments, Smith and Kida (1991) conclude that heuristics and biases documented in the psychology literature are much less common among auditors performing expert tasks. Russo et al. (2000) find that audit seniors making a nonexpert (but familiar) judgment distort information significantly less than do pharmaceutical salespersons making analogous decisions in their own field. Given the smaller (albeit significant) amount of distortion those

² Rich et al. (1997b, 492) suggest that subordinates may stylize their initial interpretation of evidence to persuade supervisors of the correctness of the subordinates' conclusions. This type of conscious, strategic behavior is different from unconscious predecisional distortion of information (Russo et al. 2000). Whether auditors anticipate both deliberate persuasion tactics and unconscious predecisional-distortion effects is the focus of Experiment 2.

audit seniors exhibited, it is an open empirical question whether auditors making judgments that require their expertise distort evidence.³

In Experiment 1, I examine whether auditors making a going-concern judgment interpret individual evidence items to support a supervisor's known views. Participants learn that the engagement partner is concerned that prior auditors were either unduly optimistic or unduly pessimistic when making going-concern judgments. For each evidence item, participants provide a score indicating the going-concern implication of that particular item. Based on the foregoing arguments, I test the following alternative hypothesis:

- H1:** Auditors who know *before* they evaluate evidence that the partner is concerned about undue pessimism (optimism) evaluate evidence more favorably (unfavorably) than auditors who do not learn the partner's concern until *after* they evaluate evidence.

Distortion's Effect on Judgments

In real-time audit review, subordinates become aware of supervisors' views more often and earlier in the audit process than they would under the traditional, sequential audit-review approach (Rich et al. 1997a). Earlier knowledge of supervisors' views provides an opportunity for subordinates to predecisionally distort evidence toward those views. If subordinates distort evidence, then their final judgments may be more consistent with supervisors' views than if subordinates had conducted a more independent initial evaluation of evidence before learning supervisors' views.

Interestingly, prior studies that demonstrate predecisional distortion in nonexpert decision settings provide no evidence that distortion actually influences judgments. In fact, the experimental designs used in those studies allow multiple explanations for why manipulated preferences might influence final judgments. For example, Russo et al. (1996) manipulate preference for one of two restaurants. Participants then evaluate attributes and choose a restaurant. Of the 26 participants, 24 choose the initially preferred restaurant. One explanation for this result is that participants unknowingly distort attribute information toward the preferred restaurant, and that this distorted information leads them to choose the preferred restaurant. Another explanation (not necessarily mutually exclusive) is that participants deliberately choose the preferred restaurant using a go-along-to-get-along heuristic (Chen et al. 1996; see also Lerner and Tetlock 1999), and that they later evaluate attributes in a way that appears to justify their choices.⁴ Determining whether predecisionally distorted information actually influences final judgments is important because unconscious distortion (Russo et al. 2000) may cause subordinates to agree more with supervisors' views than either they or their supervisors realize.

To demonstrate how accountability to audiences with *unknown* views can affect cognition, researchers frequently employ a timing manipulation wherein participants learn, either before or after they are exposed to information, that they may have to justify their

³ One reason to expect no distortion for expert judgments is that Rules 101 and 102 of the AICPA's *Code of Professional Conduct* charge auditors to be unbiased and impartial in the "performance of audit tests, the evaluation of the results, and the issuance of the audit report" (Arens and Loebbecke 2000, 86) (emphasis added). This standard of objectivity applies not only to the client-auditor relationship, but also to the subordinate-supervisor relationship: "Audit staff members should not subordinate their judgment to supervisors on the audit engagement. Staff auditors are responsible for their own judgments" (Arens and Loebbecke 2000, 92).

⁴ Although not intended to examine predecisional distortion effects, Peecher (1996) and Hackenbrack and Nelson (1996) provide related examples in accounting settings where both a predecisional-distortion effect and a deliberate choice to agree with supervisors (or clients) can explain subordinates' final choices.

decisions to an audience with unknown views. As Tetlock (1983, 287) explains, the objective of such a manipulation is:

to determine when in the cognitive processing sequence a given effect for an independent variable occurs. Does the effect occur as information is initially encoded or later at the retrieval or response selection stages? A common method of resolving this issue is to introduce the independent variable prior to stimulus exposure for half the subjects and after stimulus exposure for the other half.

Tetlock (1983), Tetlock and Kim (1987), and Kennedy (1993) use such a timing manipulation to demonstrate that the more-effortful information processing induced by accountability to an audience with unknown views leads to less-biased judgments. However, research examining accountability to audiences with *known* views has not employed such a manipulation. Instead, this research tends to presume that decision makers simply agree with those known views (Lerner and Tetlock 1999, 256).

I extend this research by manipulating the time when subordinates learn a supervisor's views.⁵ Participating auditors learn, either before or after they evaluate evidence, that the engagement partner is concerned about undue optimism or undue pessimism. Because all auditors learn the partner's view before making their going-concern judgments, they all have opportunity to make judgments that deliberately agree with the partner's view. However, auditors in the *before* condition have an additional opportunity to predecisionally distort evidence toward the partner's view. Thus, a higher tendency for auditors' going-concern judgments to agree with the partner's view in the *before* condition than in the *after* condition would be attributable to predecisional distortion of evidence (see Figure 1). I test the following alternative hypothesis:

H2: Auditors who learn *before* they evaluate evidence that the partner is concerned about undue pessimism (optimism) judge the client more (less) likely to continue as a going concern than auditors who read the partner's concern *after* evaluating evidence.

III. EXPERIMENT 1: THE EFFECTS OF PREDECISIONAL DISTORTION

Method

Participants

Sixty audit managers attending a Big 5 firm's national training course participated in Experiment 1. On average, these auditors had 5.1 years of experience (standard deviation 1.6) in the assurance department of a CPA firm. Forty (67 percent) of the auditors reported they had been involved in an engagement in which the audit team seriously considered issuing a qualified opinion due to going-concern issues.⁶

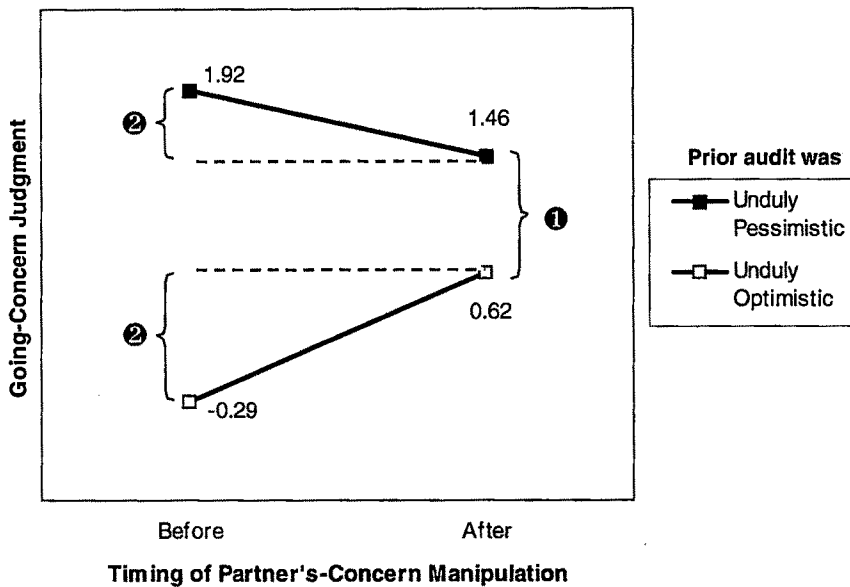
Materials

Auditors read and evaluated 20 evidence items (shown in Table 1) related to a client's going-concern status, and then made a going-concern judgment. Auditors evaluated evidence items individually on a 15-point bipolar scale with the endpoints labeled "extremely

⁵ Babcock et al. (1995) use a similar manipulation to examine self-serving estimates of settlement offers in a litigation setting. However, their manipulation allows information-search behavior, evidence weighting, and evidence evaluation to affect final judgments, whereas my manipulation distinguishes distortion from these other effects.

⁶ Neither experience nor previous involvement in going-concern deliberations differed across treatments or affected the inferences drawn from this study.

FIGURE 1
Experiment 1: Distortion's Effect on Participants' Going-Concern Judgments^a



^a This figure depicts the cell means of participants' going-concern judgments in Experiment 1. These cell means serve as the basis for planned contrasts presented in Table 3. The difference in judgments in the *after* condition (labeled with a ①) represents a deliberate or strategic agreement with the partner's view. Beyond that effect, predecisional distortion of evidence, which occurs in the *before* condition, leads to an additional difference in judgments (labeled with a ②).

NEGATIVE [POSITIVE] going-concern implications" and the center point labeled "neutral." I used these measures in testing H1. Auditors also made going-concern judgments on a 15-point bipolar scale with endpoints labeled "certain to FAIL to continue as a going concern" and "certain to CONTINUE as a going concern." The center point was labeled "completely uncertain." I used this measure in testing H2. The labeling on both scales followed Tan (1995), with slight modifications suggested by the participating firm.

I adapted evidence items from Ricchiute (1997). Ten M.B.A. students with prior audit experience and three audit managers from the participating firm rated the going-concern implications of Ricchiute's (1997) 60 evidence items. Based on these ratings, I dropped 15 items judged irrelevant to the going-concern judgment, and 8 items that were unclear to pre-test participants. From the remaining 37 items, I selected 20 items to achieve a rough balance between positive and negative going-concern implications.

Design

This experiment employed a 2×2 between-subjects factorial design. The first variable was the partner's concern. Auditors assigned to the unduly optimistic [pessimistic] condition read the following:

TABLE 1
Going-Concern Evidence Items

<i>Evidence</i>	<i>Rating^a</i>	<i>Std. Dev.</i>
1. The client defaulted on an interest payment for a bank loan.	-3.42	1.58
2. The client was denied credit from a long-time copper supplier.	-2.96	2.20
3. The client will report a net loss for the second year in a row.	-2.50	1.56
4. A principal customer has not yet placed a large expected order.	-2.00	1.60
5. Domestic and worldwide demand for wire and cable products is down.	-1.92	2.12
6. The current ratio is the lowest in the wire-and-cable industry this year.	-1.65	2.51
7. The replacement cost of assets is not fully insured.	-0.88	1.45
8. The client plans a 50 percent reduction in research costs.	-0.69	1.72
9. The client operates at 70 percent of capacity.	-0.65	1.62
10. Two wire-and-cable competitors reported net losses last year.	-0.65	1.44
11. Long-term commercial bank loan covenants restrict some fixed asset disposals.	0.26	1.69
12. The client contemplates issuing new convertible bonds next year.	0.42	1.53
13. Confirmations reveal no unrecorded indebtedness to financial institutions.	0.73	1.46
14. The client renegotiated a long-term commercial bank loan.	1.30	2.24
15. A lawyer's letter indicates that there is no current litigation.	1.50	1.77
16. The client has complied with all agreements that are contractual.	1.80	1.33
17. A newspaper reported that the client's production workers voted not to strike.	1.88	1.63
18. The client has no long-term bond indebtedness outstanding this year.	2.38	2.00
19. No trade-customer receivable accounts are 90 days or older.	2.76	2.07
20. Management signed a long-term contract with a new domestic customer.	3.30	1.26

^a Twenty-six participants who had not yet learned the partner's view indicated on a 15-point scale the implications of each individual evidence item for the client's ability to continue as a going concern (-7 = extremely NEGATIVE going-concern implications to 7 = extremely POSITIVE going-concern implications, center point labeled "neutral"). The Pearson correlation between these ratings and those of pretest participants (three managers from the participating firm and nine M.B.A. or Ph.D. students with prior audit experience) was highly significant ($r = 0.955$, $p = 0.0001$).

While helping to construct this case, the engagement partner on this audit expressed numerous times his concern that audit team members in both prior years had been *insufficiently* [overly] sensitive to evidence suggesting going-concern issues. This had led to potentially costly increases in liability exposure [unjustified investigations] and to judgments that were unduly optimistic [pessimistic].

After reading this statement, auditors listed three reasons that auditors can sometimes be unduly optimistic [pessimistic] when evaluating a client's ability to continue as a going concern. Next, auditors ranked these reasons in the order of their impact on auditors' judgments. I designed these steps to heighten the auditors' awareness of the partner's concern (e.g., Peecher 1996). The second variable was the time when auditors encountered the partner's concern. Auditors completed this manipulation either immediately before they evaluated evidence items or immediately after they evaluated evidence items.

Procedures

A partner or senior manager of the participating firm introduced the experiment and conveyed the firm's support of this research. Auditors then began Part 1 of the case by reviewing the task requirements, familiarizing themselves with the bipolar scale used to evaluate each evidence item, and then reviewing background information about the client and the results of audit procedures to date. Auditors assigned to the *before* condition completed the partner's-concern manipulation and then evaluated the 20 evidence items. Auditors assigned to the *after* condition first evaluated the 20 evidence items, then read instructions *not* to refer back to or change any of the evidence evaluations, and then completed the partner's-concern manipulation.⁷ After completing Part 1 of the case, auditors replaced it in its envelope and went on to Part 2. This part of the case began with a three-minute distracter task consisting of background questions unrelated to my hypotheses (e.g., educational level, experience). Following this task, auditors made their going-concern judgments, free-recalled evidence items, and then completed a debriefing questionnaire that contained a check for the partner's-concern manipulation. The experiment took approximately 30 minutes.

Results

Manipulation Checks

I dropped data from two auditors who failed to complete the going-concern judgment. Data from the remaining 58 auditors indicate that the manipulation of the partner's concern was successful. The results reported below are for the 53 auditors who correctly remembered the partner's concern when asked on the debriefing questionnaire. (Including the remaining five auditors who failed the manipulation check does not affect the study's inference.)

Tests of Hypotheses

I analyze the auditors' average evaluations of the 20 evidence items using a general linear model (GLM) with the following explanatory variables and their interactions: (1) the partner's concern, (2) the time when the participant learned the partner's concern, (3) the ordering of evidence items, and (4) the room to which course coordinators assigned auditors.⁸ The analysis reported in Table 2 reveals a significant interaction between the partner's concern and time ($p = 0.025$), supporting H1.⁹ Auditors in the *before* condition interpret evidence more favorably (unfavorably) when the partner is concerned about pessimism (optimism) than do auditors in the *after* condition. Simple contrasts indicate that the partner's concern has a significant effect on evidence evaluations in the *before* condition

⁷ Although in practice auditors would have the opportunity to review evidence after learning the partner's concern, and potentially would be able to reinterpret the evidence, auditors must frequently rely on their memories of evidence when making such judgments. As Libby and Trotman (1993, 561) argue, "Efficiency issues will often lead the auditor to rely on long-term memory as opposed to a thorough re-reading of the workpapers." Recall data collected later in this experiment indicate that auditors in the *before* and *after* conditions do not differ in the number of evidence items they recall ($t[51] = 0.27$, $p = 0.79$), suggesting that both groups of auditors have access to similar amounts of evidence in memory.

⁸ I crossed four random orderings of evidence with the partner's-concern and timing variables, and then randomly assigned the resulting combinations to auditors in each of three rooms. Blocking on room controlled for differences across rooms that were unrelated to my variables of interest. This was important because course coordinators had assigned auditors to rooms based on their industry specializations. Dropping room from the analysis of H1 does not change my inferences. However, dropping room from the analysis of H2 leads to an insignificant concern-by-time interaction ($p = 0.295$). Following Kuehl (1994), I report results that include room as a factor.

⁹ This concern-by-time interaction does not interact with any other variables in the model.

TABLE 2
How Partner's Concern and Time When Participants Learn that Concern Affect Participants' Evidence Evaluations—Hypothesis 1 of Experiment 1^a

Panel A: Mean Evidence Evaluations^b (Standard Deviation) [n]

<i>Partner's Concern</i>	<i>When Participants Learned the Partner's Concern</i>		
	<i>Before Evidence Evaluation</i>	<i>After Evidence Evaluation</i>	<i>Combined</i>
Prior audit was	0.10	-0.05	0.03
Unduly	(0.94)	(0.47)	(0.73)
Pessimistic	[13]	[13]	[26]
Prior audit was	-0.90	-0.05	-0.49
Unduly	(1.37)	(0.58)	(1.13)
Optimistic	[14]	[13]	[27]

Panel B: Planned Contrasts

	<i>Test Statistic</i>	<i>p-value^c</i>
A. Interaction of partner's concern with the time when participants learned that concern: $(\mu_{pb} - \mu_{ob}) > (\mu_{pa} - \mu_{oa})$	$F_{1,16} = 4.52$	0.025
B. Simple effect of the partner's concern in the <i>before</i> condition: $\mu_{pb} > \mu_{ob}$	$F_{1,16} = 12.01$	0.002
C. Simple effect of the partner's concern in the <i>after</i> condition: $\mu_{pa} > \mu_{oa}$	$F_{1,16} = 0.00$	0.490

^a Descriptive statistics and planned contrasts for participants' evidence evaluation measures. μ represents the cell mean for the partner's concern (p = pessimistic, o = optimistic) and timing (b = before, a = after) treatment combinations.

^b Participants indicated on a 15-point scale the implications of each individual evidence item for the client's ability to continue as a going concern (-7 = extremely NEGATIVE going-concern implications to 7 = extremely POSITIVE going-concern implications, center point labeled "neutral").

^c Reported p-values are one-tailed.

($p = 0.002$), but not in the *after* condition ($p = 0.490$), which makes sense given that auditors in the *after* condition had not yet learned the partner's concern.

Much of the significance in this analysis appears attributable to the *before-unduly-optimistic* cell. The partner's warning of undue optimism influenced auditors' interpretation of evidence much more than the partner's warning of undue pessimism. Although this result is consistent with prior evidence that auditors react more strongly to negative information (e.g., Smith and Kida 1991; Libby and Trotman 1993), I conduct an additional analysis on the ranked evaluation measures to ensure that the results are not driven by outliers in the *before-unduly-optimistic* cell. This ranked analysis reveals an even more significant interaction between the partner's concern and time ($p = 0.003$), again supporting H1.

To test H2, I analyze auditors' going-concern judgments using a GLM with the same explanatory variables outlined above. The analysis reported in Table 3 and depicted in Figure 1 reveals a significant interaction between the partner's-concern and time (p

TABLE 3
How Partner's Concern and Time When Participants Learn that Concern Affect Participants' Going-Concern Judgments—Hypothesis 2 of Experiment 1^a

Panel A: Mean Going-Concern Judgments^b (Standard Deviation) [n]

<i>Partner's Concern</i>	<i>When Participants Learned the Partner's Concern</i>		
	<i>Before Evidence Evaluation</i>	<i>After Evidence Evaluation</i>	<i>Combined</i>
Prior audit was	1.92	1.46	1.69
Unduly	(2.22)	(3.02)	(2.60)
Pessimistic	[13]	[13]	[26]
Prior audit was	-0.29	0.62	0.15
Unduly	(3.47)	(2.10)	(2.88)
Optimistic	[14]	[13]	[27]

Panel B: Planned Contrasts

	<i>Test Statistic</i>	<i>p-value^c</i>
A. Interaction of partner's concern with the time when participants learned that concern: $(\mu_{pb} - \mu_{ob}) > (\mu_{pa} - \mu_{oa})$	$F_{1,16} = 5.97$	0.013
B. Simple effect of the partner's concern in the <i>before</i> condition: $\mu_{pb} > \mu_{ob}$	$F_{1,16} = 6.44$	0.011
C. Simple effect of the partner's concern in the <i>after</i> condition: $\mu_{pa} > \mu_{oa}$	$F_{1,16} = 0.91$	0.177

^a Descriptive statistics and planned contrasts for participants' going-concern judgments. μ represents the cell mean for the partner's concern (p = pessimistic, o = optimistic) and timing (b = before, a = after) treatment combinations.

^b Participants indicated on a 15-point scale the likelihood that the audit client would continue as a going concern for the next year (-7 = certain to FAIL, 0 = completely uncertain, 7 = certain to CONTINUE).

^c Reported p-values are one-tailed.

= 0.013), supporting H2.¹⁰ Auditors in the *before* condition make going-concern judgments that are more consistent with the partner's concern than do auditors in the *after* condition. Simple contrasts indicate that the partner's concern has a significant effect on going-concern judgments only in the *before* condition ($p = 0.011$), not in the *after* condition ($p = 0.177$). Support for both H1 and H2 suggests that auditors making expert judgments do predecisionally distort evidence, and that this distorted evidence makes auditors' judgments more consistent with supervisors' views than deliberate or strategic agreement alone would do.

Additional Analysis

To examine the relation between predecisional distortion and memory, I also ask participants to free-recall evidence items immediately after making their going-concern judgments. I scored evidence items using a lenient gist criterion, considering items recalled

¹⁰ This concern-by-time interaction does not interact with any other variables in the model.

correct if they capture the substantial meaning of the original items (e.g., Libby and Trotman 1993; Tan 1995). On average, participants recall 7.5 of the 20 evidence items.¹¹ Based on the average evidence evaluations of the 26 managers in the *after* condition, Table 1 shows that ten evidence items are rated as positive and ten as negative. I compute net recall by subtracting the number of negative items from the number of positive items each participant recalled (e.g., Libby and Trotman 1993; Tan 1995).¹² I expect managers who know *before* evaluating evidence that the partner is concerned about undue pessimism (optimism) to have more positive (negative) net recall than managers who learn the partner's concern *after* evaluating evidence.

Using a GLM with the same explanatory variables used for H1 and H2, I observe no relation between net recall and my treatment variables. However, I do observe a positive relation between net recall and managers' going-concern judgments ($r = 0.20$, $p = 0.061$), consistent with Libby and Trotman (1993) and Tan (1995). After controlling for this relation, I observe a positive relation between net recall and managers' evidence evaluation measures ($r = 0.37$, $p = 0.003$). As managers' evidence evaluations become more favorable, they recall relatively more positive items than negative items. This is the first demonstration that predecisional distortion of evidence is related to memory, and suggests that the impact of predecisional distortion is likely to extend beyond judgments and decisions.

IV. EXPERIMENT 2: ANTICIPATING PREDECISIONAL DISTORTION

In real-time audit review, subordinates are likely to learn supervisors' views earlier in the audit process. If subordinates then unconsciously distort evidence toward supervisors' views, audit conclusions are likely to be adversely affected. If supervisors do not anticipate this increased tendency for subordinates to agree, or cannot detect it because documented evidence largely supports subordinates' judgments (Libby and Trotman 1993; Koonce et al. 1995; Rich 1998; Ricchiute 1999), then audit errors are more likely to occur. Prior research suggests that supervisors' knowledge of their own persuasion tactics is likely to help them anticipate a subordinate's deliberate or strategic agreement with their views (Rich et al. 1997b). However, supervisors are unlikely to have learned to anticipate predecisional-distortion effects because distortion itself is largely unconscious (Russo et al. 2000). In the second experiment, I examine whether auditors anticipate the effects, whether deliberate or unconscious, that supervisors' views can have on subordinates' judgments.

Prior research suggests that auditors are often too partial to their own views (e.g., Tan 1995), a tendency that is likely to hinder auditors' ability to anticipate predecisional distortion toward their own views. To avoid biasing my inferences toward concluding that auditors are unable to anticipate predecisional distortion, I do not ask auditors to express their own views in Experiment 2. Instead, I ask an independent group of auditors to anticipate the going-concern judgment a new manager would make in response to an engagement partner's view, given the time when the manager learned this view. In a between-subjects design, I manipulate the partner's view and the time when the manager learned the partner's view. If auditors' knowledge of persuasion tactics helps them anticipate subordinates' deliberate or strategic agreement with supervisors' views (Rich et al. 1997b), then auditors

¹¹ This number does not vary significantly across treatments, and controlling for it does not influence any of the previously reported results.

¹² The results reported here do not change when I ignore the middle two, four, six, or eight evidence items when defining whether an item is positive or negative.

will expect the manager's going-concern judgment to be consistent with the partner's views. This suggests the following alternative hypothesis:

H3: When the partner is concerned about undue pessimism rather than undue optimism, auditors expect managers to judge the client *more* likely to continue as a going concern.

On the other hand, if the unconscious nature of predecisional distortion has kept auditors from becoming aware of distortion effects (Russo et al. 2000), then auditors are unlikely to anticipate how earlier knowledge of the partner's view increases the manager's tendency to agree with that view. I therefore expect *not* to reject the following null hypothesis:

H4: Auditors do *not* expect the partner's concern to have more influence on the manager's going-concern judgments when the manager learns of the partner's concern *before* evaluating evidence rather than *after* evaluating evidence.

By not asking auditors to express their own views and then anticipate how a subordinate would respond, my experimental design works in favor of rejecting both H3 and H4. This bias is in favor of my expectations for H3, but against my expectations for the more important H4, which posits that auditors do not anticipate predecisional distortion, and therefore do not expect the partner's concern to have more influence on the manager's judgment when the manager learns of the partner's concern *before* evaluating evidence. If auditors do not anticipate predecisional-distortion toward *others'* views (as in my experimental setting), then it is even less likely that auditors would anticipate predecisional distortion toward their *own* views.

Method

Participants

Because Experiment 2 asks participants to anticipate how a new manager would respond to a partner's view, the ideal participants would be partners themselves. However, the cost of partners' time led me to use less experienced supervising seniors. Supervising seniors spend considerable time reviewing the work of junior team members, so they should be familiar with how subordinates respond to supervisors' views. It is this familiarity with or understanding of subordinate-supervisor interactions that Experiment 2 examines. Moreover, given that distortion is hard to detect in one's own judgment (Russo et al. 2000), and given the current lack of firm training on this topic, it is unlikely that a partner would be more aware of the unconscious effects of distortion than would be a supervising senior.¹³

One hundred twenty supervising seniors attending a Big 5 firm's national training course participated in this experiment. On average, these auditors had 3.4 years of experience (standard deviation 1.4) in the assurance practice of a CPA firm. Eighty (67 percent)

¹³ I considered developing an alternative scenario in which supervising seniors predict responses of *less* experienced auditors (i.e., junior staff). Unfortunately, if supervising seniors in this new scenario did not anticipate the deliberate and unconscious effects of supervisors' views on subordinates' judgments, then it could be because either (1) supervising seniors do not understand these effects, or (2) supervisors' views simply have no effect on subordinates' judgments in the new scenario. To avoid this second potential explanation, I decided to retain the manager/partner scenario from Experiment 1.

of the auditors reported they had been involved in an engagement in which the audit team seriously considered issuing a qualified opinion due to going-concern issues.¹⁴

Design

The experiment employed a 2×2 between-subjects factorial design. The first variable was the partner's concern, which I manipulated as in Experiment 1. The second variable was the time when the new manager became aware of the partner's concern. Participants were told that the engagement partner held a meeting with the new manager to discuss overall audit-planning issues. Participants in the *before* (*after*) condition read that this meeting occurred *before* (*after*) the manager evaluated evidence related to client going-concern issues. The partner's-concern manipulation and the timing-description manipulation appeared on the same page and were counterbalanced to control for potential order effects. I randomly assigned the four treatment combinations of this variable to participants in each of four rooms.

Procedures and Materials

A partner of the participating firm introduced the experiment and conveyed the firm's support of this research. Participants then began Part 1 of the case by reading the instructions and familiarizing themselves with the bipolar scale used to anticipate the new manager's going-concern judgment. Next, they reviewed background information about the client and the results of audit procedures to date, and then reviewed the 20 evidence items from Experiment 1. Next, participants encountered the partner's-concern and timing-description manipulations. Finally, they indicated the going-concern judgment they thought the new manager would make using a similar scale to that in Experiment 1. In Part 2 of the case, participants completed a manipulation check for the partner's concern and timing variables. Next, they anticipated the going-concern judgment that the new manager would make in the opposite timing condition to what they had encountered in Part 1. Finally, they completed a debriefing questionnaire. The entire experiment took approximately 30 minutes.

Results

Manipulation Checks

The results are based on 111 (out of the 120) participating auditors who remembered the partner's concern and the time of the partner-manager meeting when asked during debriefing. Including the responses of the remaining nine auditors who failed one of the two manipulation checks does not affect the study's inferences.

Tests of Hypotheses

I analyzed participants' anticipated going-concern judgments using a GLM with the following explanatory variables and their interactions: (1) the partner's concern, (2) the time when the manager learned about this concern, (3) the room to which course coordinators assigned participants, and (4) the order in which I presented the partner's-concern and timing-description variables. Results reported in Table 4 and depicted in Panel A of Figure 2 reveal a significant main effect for the partner's concern ($p = 0.022$), supporting H3.¹⁵ Participants expect managers to make judgments that agree with the partner's concern. However, the analysis reveals an insignificant interaction between the partner's-concern and

¹⁴ Neither experience nor previous involvement in going-concern deliberations differed across treatments or affected the inferences drawn from this study.

¹⁵ The partner's-concern variable does not interact with any other variables in the model.

timing-description variables ($p = 0.381$), suggesting that participants do not expect the partner's concern to have any more influence on managers' going-concern judgments in the *before* condition than in the *after* condition. Thus, as expected, I fail to reject the null hypothesis expressed in H4. These results suggest that participants anticipate subordinates' deliberate agreement with supervisors' views, but they do not anticipate how earlier knowledge of supervisors' views increases subordinates' tendency to agree with those views.

Corrective Intervention

To examine a potential corrective intervention, I asked participants in a follow-up question to anticipate what the manager's going-concern judgment would be in the opposite

TABLE 4
Anticipating Distortion's Effect on Subordinates' Going-Concern Judgments—Hypotheses 3 and 4 of Experiment 2^a

Panel A: Mean Anticipated Going-Concern Judgments^b (Standard Deviation) [*n*]

Partner's Concern	Timing of Partner-Manager Meeting		Combined
	Before Manager Evaluated Evidence	After Manager Evaluated Evidence	
Prior audit was	1.93	1.67	1.80
Unduly	(2.69)	(2.76)	(2.70)
Pessimistic	[27]	[27]	[54]
Prior audit was	0.63	0.68	0.65
Unduly	(2.70)	(3.24)	(2.92)
Optimistic	[32]	[25]	[57]

Panel B: Planned Contrasts

	Test Statistic	<i>p</i> -value ^c
A. Interaction of partner's concern with the time when manager learned that concern: $(\mu_{pb} - \mu_{ob}) > (\mu_{pa} - \mu_{oa})$	$F_{1,80} = 0.09$	0.381
B. Simple effect of the partner's concern in the <i>before</i> condition: $\mu_{pb} > \mu_{ob}$	$F_{1,80} = 3.05$	0.042
C. Simple effect of the partner's concern in the <i>after</i> condition: $\mu_{pa} > \mu_{oa}$	$F_{1,80} = 1.56$	0.108
D. Main effect of the partner's concern: $(\mu_{pb} + \mu_{pa}) > (\mu_{ob} + \mu_{oa})$	$F_{1,80} = 4.19$	0.022

^a Descriptive statistics and planned contrasts for managers' going-concern judgments as anticipated by participating auditors. μ represents the cell mean for the partner's concern (p = pessimistic, o = optimistic) and timing (b = before, a = after) treatment combinations.

^b Participants indicate on a 15-point scale the likelihood the new manager would conclude that the audit client would continue as a going concern for the next year (-7 = certain to FAIL, 0 = completely uncertain, 7 = certain to REMAIN VIABLE).

^c All reported *p*-values are one-tailed.

timing-description condition. That is, I asked participants in the *before* condition to anticipate the manager's judgment if the partner had expressed his view *after* the manager evaluated evidence, and vice versa. This question highlighted how the time at which guidance is given can vary. I analyzed responses to this question using a GLM with the same explanatory variables outlined above. Results appear in Table 5 and in Panel B of Figure 2. Here, the analysis reveals a significant interaction between the partner's-concern and timing-description variables ($p = 0.047$), indicating that participants expect the partner's concern to have a more significant effect on going-concern judgments in the *before* condition ($p = 0.008$) than in the *after* condition ($p = 0.247$).¹⁶ This result suggests that simply asking participants to focus on the time at which guidance is given is enough to trigger their anticipation of subordinates' increased tendency to agree. This, in turn, suggests that training is likely to help auditors anticipate the predecisional-distortion effects supervisors' views can have on subordinates' judgments.

V. DISCUSSION

This study's results suggest that earlier knowledge of supervisors' views influences subordinates' judgments, in part because subordinates predecisionally distort evidence toward those views. Results also suggest that auditors likely do not anticipate how earlier knowledge of supervisors' views increases subordinates' tendency to agree with those views. Together, these results suggest that the current trend toward real-time audit review, with its earlier and more frequent guidance from supervisors, is likely to have unexpected consequences that could lead to audit errors. These findings suggest that corrective interventions should focus on how subordinates evaluate evidence or on supervisors' ability to anticipate both the deliberate and the unconscious effects their views can have on subordinates' judgments.

The existence of predecisional distortion has implications beyond the audit-review setting. For example, Plumlee (1985) demonstrates that auditors who participate in the creation of an internal control system later attribute less control risk to that system than auditors who do not participate in its creation. Although Plumlee (1985) uses internal auditors, the results speak to situations in which external auditors assist their clients in establishing procedures (i.e., through management advisory services), and are later asked to evaluate the strength of those systems. One reason that these auditors attribute less control risk to the systems they help create may be due to predecisional distortion of evidence. That is, auditors may interpret control-system evidence to appear consistent with lower control risk. Future research might examine the extent to which predecisional distortion drives auditors' lack of independence as documented in Plumlee (1985).

Predecisional distortion of information also provides a potential explanation for decision-maker overconfidence. Russo et al. (1998) demonstrate that confidence and predecisional distortion are positively related. Decision makers may become overconfident by interpreting new evidence consistently with their currently held beliefs or preferences. Although a number of studies have identified what some consider an overconfidence phenomenon in market settings (e.g., De Bondt and Thaler 1985; Bloomfield et al. 2000), only recently have studies begun to examine the underlying psychological processes that may lead to such overconfidence. Future research could examine the extent to which persistent overconfidence is due to predecisional distortion of new information. Such a relation would

¹⁶ This concern-by-time interaction does not interact with any other variables in the model.

TABLE 5
Can Corrective Intervention Help Participants Anticipate Distortion's Effect on Subordinates' Going-Concern Judgments?—Experiment 2^a

Panel A: Mean Anticipated Going-Concern Judgments^b (Standard Deviation) [n]

<i>Partner's Concern</i>	<i>Timing of Partner-Manager Meeting</i>		<i>Combined</i>
	<i>Before Manager Evaluated Evidence</i>	<i>After Manager Evaluated Evidence</i>	
Prior audit was	1.52	0.48	1.00
Unduly	(3.54)	(3.33)	(3.45)
Pessimistic	[27]	[27]	[54]
Prior audit was	-0.92	-0.16	-0.49
Unduly	(3.63)	(3.24)	(3.41)
Optimistic	[25]	[32]	[57]

Panel B: Planned Contrasts

	<i>Test Statistic</i>	<i>p-value^c</i>
A. Interaction of partner's concern with the time when manager learned that concern: $(\mu_{pb} - \mu_{ob}) > (\mu_{pa} - \mu_{oa})$	$F_{1,80} = 2.88$	0.047
B. Simple effect of the partner's concern in the <i>before</i> condition: $\mu_{pb} > \mu_{ob}$	$F_{1,80} = 6.10$	0.008
C. Simple effect of the partner's concern in the <i>after</i> condition: $\mu_{pa} > \mu_{oa}$	$F_{1,80} = 0.47$	0.247
D. Main effect of the partner's concern: $(\mu_{pb} + \mu_{pa}) > (\mu_{ob} + \mu_{oa})$	$F_{1,80} = 6.41$	0.007

^a Descriptive statistics and planned contrasts for managers' going-concern judgments as anticipated by participating auditors who had been exposed to corrective intervention. μ represents the cell mean for the partner's concern (p = pessimistic, o = optimistic) and timing (b = before, a = after) treatment combinations.

^b Participants indicate on a 15-point scale the likelihood the new manager would conclude that the audit client would continue as a going concern for the next year (-7 = certain to FAIL, 0 = completely uncertain, 7 = certain to REMAIN VIABLE).

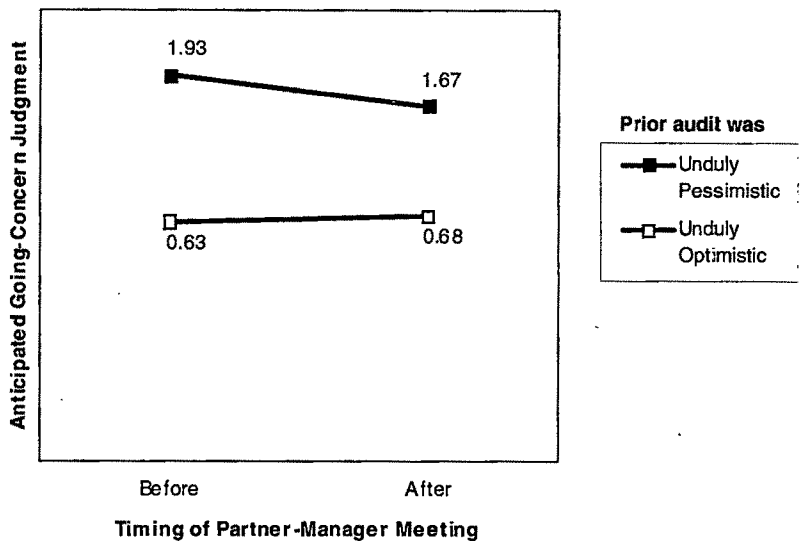
^c All reported p-values are one-tailed.

suggest that giving more information to less-informed but overconfident investors may exacerbate overconfidence (cf. Bloomfield et al. 1999).

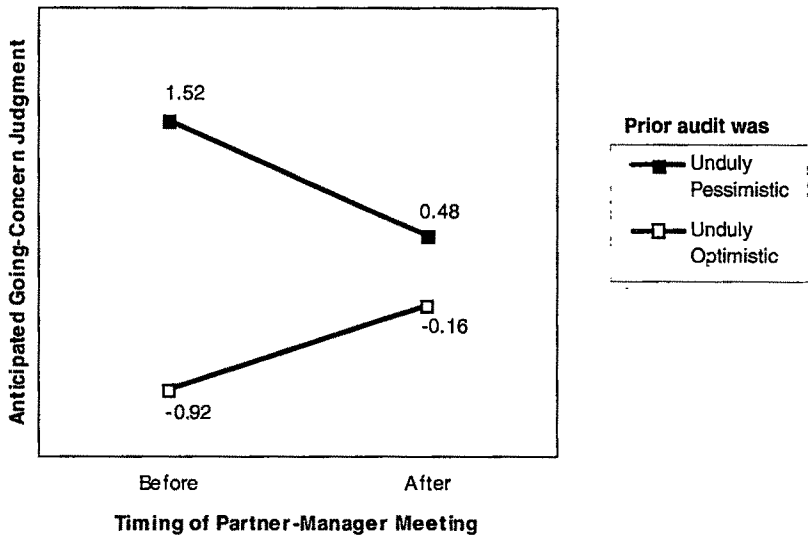
This study's results extend the motivated-reasoning literature by demonstrating the incremental impact that evidence evaluation (i.e., predecisional distortion) has on subordinates' judgments. However, these results may understate the total effect supervisors' views can have on subordinates' judgments because those views also influence processes such as hypothesis generation and evidence search. For example, Cloyd and Spilker (1999) show that tax professionals spend more time viewing cases with conclusions that are consistent with their clients' desired outcomes. Viewing time then predicts participants' likelihood

FIGURE 2
Experiment 2: Participants' Anticipation of Distortion Effects^a

Panel A: Participants' Anticipated Judgments (before Corrective Intervention)



Panel B: Participants' Anticipated Judgments (after Corrective Intervention)



^a Panel A of this figure depicts participants' *anticipated* going-concern judgments in Experiment 2. Panel B depicts the *anticipated* going-concern judgments participants made after being exposed to a corrective intervention that highlighted how the timing of the partner-manager meeting could vary.

assessments that a court would rule in favor of the client's position. These assessments then predict participants' final recommendations. If audit subordinates behave similarly by searching for evidence that confirms their supervisors' views, then not only will their evaluations of evidence be distorted as demonstrated in this study, but they will also perform these evaluations on a set of evidence already biased toward the supervisors' views.

Limitations

In Experiment 1, auditors in the *after* condition could not review evidence after learning the partner's concern. Although auditors often must rely on memory of evidence rather than on a thorough rereading of working papers (Libby and Trotman 1993), this aspect of my design may have kept auditors from re-interpreting information. Future research might address whether predecisional distortion of evidence still influences judgments when subordinates have the chance to review evidence just before making their final judgments. Future research might also extend our understanding of the relation between distortion and memory. To the degree that distortion influences memory, later audit decisions that rely on memory are also likely to be affected (Rau and Moser 1999).

In Experiment 1, the limited amount of information presented to participants may have made it easier for them to distort predecisionally. Although it would seem that earlier pieces might be more easily distorted, the positive relation between confidence and distortion documented in prior studies (Russo et al. 1998) suggests that later pieces may be more distorted if confidence increases during the evaluation process.

Finally, in Experiment 2, I used supervising seniors, rather than partners, to anticipate how managers would respond to a partner's guidance. Although a number of studies have found differences between senior, manager, and partner performance on dimensions such as subspecialty knowledge and related general business knowledge (Bonner and Lewis 1990), tacit managerial knowledge (Tan and Libby 1997), and the ability to predict another auditor's choice (Jamal and Tan 2001), the extent to which more experienced auditors anticipate the unconscious predecisional-distortion effects this study identifies remains unexplored. Future research examining this issue must first specify when (or provide circumstances in which) auditors can learn to anticipate such distortion effects (Libby and Luft 1993).

REFERENCES

- Arens, A., and J. Loebbecke. 2000. *Auditing. An Integrated Approach*. Englewood Cliffs, NJ: Prentice Hall.
- Babcock, L., G. Loewenstein, S. Issacharoff, and C. Camerer. 1995. Biased judgments of fairness in bargaining. *The American Economic Review* 85 (5): 1337–1343.
- Bloomfield, R., R. Libby, and M. Nelson. 1999. Confidence and the welfare of less-informed investors. *Accounting, Organizations and Society* 24 (8): 623–647.
- , ———, and ———. 2000. Under-reactions, over-reactions, and moderated confidence. *Journal of Financial Markets* 3 (May): 113–137.
- Boiney, L., J. Kennedy, and P. Nye. 1997. Instrumental bias in motivated reasoning: More when more is needed. *Organizational Behavior and Human Decision Processes* 72 (1): 1–24.
- Bonner, S., and B. Lewis. 1990. Determinants of auditor expertise. *Journal of Accounting Research* 28 (Supplement): 1–20.
- , R. Libby, and M. Nelson. 1996. Using decision aids to improve auditors' conditional probability judgments. *The Accounting Review* 71 (2): 221–240.
- Brown, C., M. Peecher, and I. Solomon. 1999. Auditors' hypothesis testing in diagnostic inference tasks. *Journal of Accounting Research* 37 (1): 1–26.

- Buchman, T., P. Tetlock, and R. Reed. 1996. Accountability and auditor's judgments about contingent events. *Journal of Business Finance and Accounting* 23 (3): 379-399.
- Chen, S., D. Shechter, and S. Chaiken. 1996. Getting at the truth or getting along: Accuracy- versus impression-motivated heuristic and systematic processing. *Journal of Personality and Social Psychology* 71 (2): 262-275.
- Church, B. K. 1991. An examination of the effect that commitment to a hypothesis has on auditors' evaluations of confirming and disconfirming evidence. *Contemporary Accounting Research* 7 (2): 513-534.
- , and A. Schneider. 1993. Auditor's generation of diagnostic hypotheses in response to a superior's suggestion: Interference effects. *Contemporary Accounting Research* 10 (1): 333-350.
- Cloyd, B., and B. Spilker. 1999. The influence of client preferences on tax professionals' search for judicial precedents, subsequent judgment, and recommendations. *The Accounting Review* 74 (3): 299-322.
- Cuccia, A., K. Hackenbrack, and M. Nelson. 1995. The ability of professional standards to mitigate aggressive reporting. *The Accounting Review* 70 (2): 227-248.
- De Bondt, W., and R. Thaler. 1985. Does the stock market overreact? *The Journal of Finance* 40 (3): 793-808.
- Hackenbrack, K., and M. Nelson. 1996. Auditors' incentives and their application of financial accounting standards. *The Accounting Review* 71 (1): 43-59.
- Heiman, V. 1990. Auditors' assessment of the likelihood of error explanations in analytical review. *The Accounting Review* 65 (4): 875-890.
- Jamal, K., and H. T. Tan. 2001. Can auditors predict the choices made by other auditors? *Journal of Accounting Research* 39 (3): 583-597.
- Johnson, L. 1993. An empirical investigation of the effects of advocacy on preparers' evaluations of judicial evidence. *The Journal of the American Taxation Association* 15 (1): 1-22.
- Kadous, K., J. Kennedy, and M. Peecher. 2001. Are auditors objective when assessing the relative quality of client accounting methods? Working paper, University of Washington, Seattle, WA.
- Kennedy, J. 1993. Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research* 31 (2): 231-245.
- . 1995. Debiasing the curse of knowledge in audit judgment. *The Accounting Review* 70 (2): 249-273.
- Koonce, L., U. Anderson, and G. Marchant. 1995. Justification of decisions in auditing. *Journal of Accounting Research* 33 (2): 369-384.
- Kuehl, R. 1994. *Statistical Principles of Research Design and Analysis*. Belmont, CA: Duxbury.
- Kunda, Z. 1990. The case for motivated reasoning. *Psychological Bulletin* 108 (3): 480-498.
- Lerner, J., and P. Tetlock. 1999. Accounting for the effects of accountability. *Psychological Bulletin* 125 (2): 255-275.
- Libby, R. 1981. *Accounting and Human Information Processing: Theory and Applications*. Englewood Cliffs, NJ: Prentice Hall.
- . 1985. Availability and the generation of hypotheses in analytical review. *Journal of Accounting Research* 23 (2): 648-667.
- , and P. Libby. 1989. Expert measurement and mechanical combination in control reliance decisions. *The Accounting Review* 64 (4): 729-747.
- , and J. Luft. 1993. Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, Organizations and Society* 18 (5): 425-450.
- , and K. Trotman. 1993. The review process as a control for differential recall of evidence in auditor judgments. *Accounting, Organizations and Society* 18 (6): 559-574.
- McMillan, J., and R. White. 1993. Auditors' belief revisions and evidence search: The effect of hypothesis frame, confirmation bias, and professional skepticism. *The Accounting Review* 68 (3): 443-465.
- Nelson, M., and W. Kinney. 1997. The effect of ambiguity on loss contingency reporting judgments. *The Accounting Review* 72 (2): 257-274.

- Peecher, M. 1996. The influence of auditors' justification processes on their decisions: A cognitive model and experimental evidence. *Journal of Accounting Research* 34 (1): 125–141.
- Plumlee, R. D. 1985. The standard of objectivity for internal auditors: Memory and bias effects. *Journal of Accounting Research* 23 (2): 683–699.
- Ponemon, L. 1995. The objectivity of accountants' litigation support judgments. *The Accounting Review* 70 (3): 467–488.
- Pyszczynski, T., and J. Greenberg. 1987. Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In *Advances in Experimental Social Psychology*, Vol. 20, edited by L. Berkowitz, 297–340. New York, NY: Academic Press.
- Rau, S., and D. Moser. 1999. Does performing other audit tasks affect going-concern judgments? *The Accounting Review* 74 (4): 493–508.
- Ricchiute, D. 1997. Effects of judgment on memory: Experiments in recognition bias and process dissociation in a professional judgment task. *Organizational Behavior and Human Decision Processes* 70 (1): 27–39.
- . 1999. The effect of audit seniors' decisions on working paper documentation and on partners' decisions. *Accounting, Organizations and Society* 24: 155–171.
- Rich, J., I. Solomon, and K. Trotman. 1997a. Multi-auditor judgment/decision making research: A decade later. *Journal of Accounting Literature* 16: 86–126.
- , ———, and ———. 1997b. The audit review process: A characterization from the persuasion perspective. *Accounting, Organizations and Society* 22: 481–505.
- . 1998. An experimental investigation of the review process from a persuasion perspective. Working paper, University of Connecticut, Storrs, CT.
- Russo, J., V. Medvec, and M. Meloy. 1996. The distortion of information during decisions. *Organizational Behavior and Human Decision Processes* 66 (1): 102–110.
- , M. Meloy, and V. Medvec. 1998. Predecisional distortion of product information. *Journal of Marketing Research* 34 (November): 438–452.
- , ———, and T. J. Wilks. 2000. Predecisional distortion of information by auditors and salespersons. *Management Science* 46 (1): 13–27.
- Salterio, S., and L. Koonce. 1997. The persuasiveness of audit evidence: The case of accounting policy decisions. *Accounting, Organizations and Society* 22 (6): 573–587.
- Smith, J. F., and T. Kida. 1991. Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin* 109 (3): 472–489.
- Tan, H. T. 1995. Effects of expectations, prior involvement, and review awareness on memory for audit evidence and judgment. *Journal of Accounting Research* 33 (1): 113–135.
- , and R. Libby. 1997. Tacit managerial knowledge versus technical knowledge as determinants of audit expertise in the field. *Journal of Accounting Research* 35 (1): 97–113.
- Tetlock, P. 1983. Accountability and the perseverance of first impressions. *Social Psychology Quarterly* 46 (4): 285–292.
- , and J. Kim. 1987. Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology* 52 (4): 700–709.
- . 1992. The impact of accountability on judgment and choice: Toward a social contingency model. In *Advances in Experimental Social Psychology*, Vol. 25, edited by L. Berkowitz, 331–376. New York, NY: Academic Press.

Revisiting the Reportedly Weak Value Relevance of Oil and Gas Asset Present Values: The Roles of Measurement Error, Model Misspecification, and Time-Period Idiosyncrasy

Jeff P. Boone
Mississippi State University

ABSTRACT: This paper investigates three potential explanations for the puzzlingly weak value relevance of oil and gas asset present values documented in prior research: measurement error, model misspecification, and time-period idiosyncrasy. I operationally define the magnitude of measurement error as the measurement error variance, estimated using an errors-in-variables two-stage regression model similar to that used by Barth (1991) and Choi et al. (1997). I find that (1) measurement error in the present value measure of oil and gas assets is on average less than the measurement error in the historical cost asset measure; (2) oil and gas assets measured at present value explain significantly more across-firm and across-time variation in stock prices than do oil and gas assets measured at historical cost; and (3) model misspecification partially accounts for the puzzlingly weak reported value relevance of the present value measure in prior research.

Keywords: *measurement error; reliability; present value; oil and gas.*

Data Availability: *Data are available from sources identified in the paper.*

I. INTRODUCTION

Accounting use of fair value measures is increasing, and with it the use of present value as an estimate of fair value. For example, 11 of the 32 financial reporting standards

I appreciate helpful comments by Mary Barth (the associate editor), Charles Boynton, and two anonymous reviewers. Funding for much of this project was provided by the Institute of Petroleum Accounting at the University of North Texas.

*Submitted February 2000
Accepted July 2001*

the Financial Accounting Standards Board (FASB) issued between December 1990 and December 1999 required firms to measure present values (FASB 2000, paragraph 4).¹ Yet despite this increase, our understanding of when and why such measures are relevant and reliable is limited. My study adds to our understanding of the relevance and reliability of present value measures by investigating three possible alternative explanations for the reportedly weak value relevance of oil and gas asset present values documented in prior research: measurement error, model misspecification, and time-period idiosyncrasy.

Firms have disclosed present value estimates of oil and gas asset fair values for two decades, pursuant to Accounting Securities Release (ASR) No. 253 and later Statement of Financial Accounting Standards (SFAS) No. 69. Prior research based on 1979–1988 data has concluded that, at best, these present value disclosures have only weak explanatory power for stock prices (e.g., Magliolo 1986; Harris and Ohlson 1987; Shaw and Wier 1993) and stock returns (Doran et al. 1988; Alciatore 1993). The apparently weak value relevance of these disclosures is puzzling because the economic character of oil and gas assets suggests that their present values should be no less value relevant than are other types of operating assets. Prior research has speculated that the weak association between these present value estimates and stock prices/returns is attributable to measurement error in the present value measure (Harris and Ohlson 1987; Magliolo 1986; Lys 1986; Clinch and Magliolo 1992). Measurement error seemed a plausible explanation because of alleged defects in the method SFAS No. 69 prescribed to assign value to reserve quantities, and because reserve quantities themselves are difficult to estimate accurately. This explanation now appears less plausible in light of recent evidence of a fairly consistent association between stock prices and fair value estimates not only of financial assets (e.g., Landsman 1986; Barth 1991, 1994; Barth et al. 1996; Eccher et al. 1996; Venkatachalam 1996; Nelson 1996), but also of long-lived operating assets with substantial firm-specific value for which fair values are difficult to estimate accurately (i.e., intangible assets and property and equipment; see Easton et al. 1993; Amir et al. 1993; Barth and Clinch 1998; Aboody et al. 1999). In contrast to these firm-specific operating assets, oil and gas assets trade in active markets; these assets have little firm-specific value, and thus their fair values should be accurately estimable. Since measurement error is insufficient to generate statistical insignificance in tests of value relevance of the hard-to-measure fair values of long-lived assets,² it seems unlikely that measurement error accounts for the generally insignificant association between stock prices and the more easily measured present value of oil and gas assets. Thus, prior research's weak evidence that present values of oil and gas assets are value relevant presents an unsolved puzzle. I test measurement error, model misspecification, and time-period idiosyncrasy as three explanations for the reportedly weak value relevance of oil and gas asset present values.

I am able to replicate the findings of prior research when estimating annual, imputed value models similar to those estimated in the studies. In particular, during 1984–88, the years common to both my study and prior studies, the present value measure of oil and gas assets exhibits significantly *less* explanatory power for stock prices than the corresponding historical cost measure. However, specification tests reject the annual, imputed value

¹ The FASB issued Statement of Financial Accounting Concepts (SFAC) No. 7 in February 2000 to provide conceptual guidance on the accounting use of present values.

² I use the term *value relevance* to mean a statistically significant association with stock prices or stock returns in the predicted direction.

model in favor of an unrestricted, fixed-effects balance-sheet valuation model. In the unrestricted, fixed-effects model, oil and gas assets measured at present value exhibit significantly *more* explanatory power for stock prices than the corresponding historical cost measure, both in the full 1984–1996 period and in the 1984–88 subperiod.

Consistent with the valuation analysis, I also find evidence that measurement error in the present value measure of oil and gas assets is *less* than in the corresponding historical cost measure, both in the 1984–1996 period and the 1984–88 subperiod that corresponds to the period examined in prior research. Analysis of across-firm and across-time variation in measurement-error variance reveals that present value measurement-error variance increases both as revisions in reserve quantity estimates increase and as firm-specific discount rates deviate from the uniform 10 percent rate SFAS No. 69 requires in calculating present values.

Overall, the evidence suggests that model misspecification, rather than measurement error or time-period idiosyncrasy, most likely explains the weak value relevance of oil and gas present values reported in prior research. These results help resolve an anomaly in the literature, and demonstrate that present value offers a viable means of obtaining reliable fair value estimates for oil and gas assets—nonfinancial assets that have little firm-specific value and are traded in active markets.

I have organized the remainder of the paper as follows: Section II provides institutional detail concerning oil and gas accounting, discusses related prior research, and develops the three hypotheses. Section III develops the research design, while Section IV discusses the sample selection criteria and characteristics of the sample. Section V discusses test results, and synthesis and conclusions follow in Section VI.

II. BACKGROUND, PRIOR RESEARCH, AND HYPOTHESES

Background and Prior Research

Firms engaged in oil and gas exploration and production can use either the full cost or the successful efforts method to account for their oil and gas activities. The two methods differ in the accounting for the cost of drilling nonproductive wells. The successful efforts method expenses these costs when the well is deemed nonproductive, whereas the full cost method capitalizes and amortizes these costs under the rationale that the cost incurred to discover producing wells necessarily includes the cost of drilling nonproductive wells. Thus, full cost accounting leads to higher historical cost of oil and gas assets than does successful efforts accounting. Because required asset impairment tests effectively limit an asset's depreciated cost to no more than current value, a higher oil and gas asset cost is closer to fair value than a lower cost.

Critics allege that full cost and successful efforts methods, both based on historical cost accounting, are inappropriate for firms in the oil and gas industry (U.S. Securities and Exchange Commission [SEC] 1978). Companies often discover oil and gas assets with substantial value at relatively nominal cost, so the cost of obtaining these assets is often not a fair measure of the value obtained at the time of acquisition. Critics argue that this violation of the fundamental premise underlying historical cost accounting renders historical cost inappropriate.

Because of this limitation, the SEC in 1978 sought to replace historical cost accounting for oil and gas activities with reserve recognition accounting (SEC 1978) based on present value rather than historical cost asset measurement. The SEC, facing stiff opposition based on concerns that the present value measure lacked reliability, eventually abandoned the proposal (SEC 1981). The FASB, however, made this present value measure the centerpiece of the supplemental disclosures required by SFAS No. 69 (FASB 1982). SFAS No. 69

requires firms to estimate oil and gas asset fair values by discounting at a 10 percent rate of interest the net revenue from estimated future extraction and sale of "proved" reserves.

Prior research concludes that present value estimates of oil and gas asset fair values prepared pursuant to SFAS No. 69 are, at best, weakly value relevant. Magliolo (1986) finds that while the present value measure of oil and gas assets is associated with stock prices during the period 1979–1983, the association is weaker than that based on a competing proprietary estimate of reserve values prepared by financial analysts. Magliolo (1986) speculates, as does Lys (1986), that the present value measure of oil and gas assets may contain substantial measurement error. Harris and Ohlson (1987) find that during the period 1979–1983, the historical cost measure of oil and gas assets is more strongly associated with stock prices than is the present value measure, leading them also to speculate that the present value measure suffers from measurement error. Using a returns-type design and data obtained from 1979–1984, Doran et al. (1988) find little association between stock returns and changes in the SFAS No. 69 measure, while Alciatore (1993) finds that selected components of the change in the SFAS No. 69 measure are associated with stock returns during the period 1982–84. Shaw and Wier (1993) find little evidence of an association between stock prices and the SFAS No. 69 present value measure during the period 1985–88. As I elaborate below, I investigate three possible explanations for the weak value relevance: measurement error, model misspecification, and time-period idiosyncrasy hypotheses.

Measurement Error Hypothesis

Barth and Landsman (1995) note that market imperfections create a divergence among asset entry value (i.e., replacement cost), exit value (i.e., the price at which an asset can be sold to a third party), and value-in-use (i.e., the incremental firm value attributable to an asset). Value-in-use differs from exit value if a firm's unique skill in exploiting the asset creates firm-specific value that is not transferable to a buyer. Barth and Landsman (1995) further note that it is particularly difficult to accurately estimate fair values for assets that are not traded in active markets and that have substantial firm-specific value. Oil and gas assets are actively traded in established markets and have little firm-specific value. Thus, present value estimates of oil and gas asset fair values should exhibit *less* measurement error than fair value estimates of intangible assets and property and equipment that are not traded in established markets and that contain substantial firm-specific value. On the other hand, the SFAS No. 69 present value measure may be subject to severe measurement error because (1) reserve quantity estimates underlying the valuation disclosures may contain substantial error, and (2) the valuation model used to attach value to those reserve quantities may be flawed (Clinch and Magliolo 1992).

Studies conclude that estimates of reserve quantities do contain substantial error. For example, the annual *revision* of the previous year's reserve quantity estimate as a percentage of beginning reserve quantities ranges from –82 percent to 37 percent with a standard deviation of 17 percent around a mean of –3.9 percent (Porter 1980; King 1982; Kahn et al. 1983). Furthermore, this annual revision of reserve estimates varies across firms and across time in a pattern suggesting that petroleum firms use reserve estimates to manage earnings (Hall and Stammerjohan 1997). Thus, noise in reserve quantity estimates, whether from estimation error or purposeful manipulation, may introduce measurement error into oil and gas asset present values. *Ceteris paribus*, firms that tend to report large annual revisions to the previous year's reserve quantity estimates should have greater present value measurement error. Moreover, present value measurement error may also depend upon whether internal or external petroleum engineers prepare the reserve estimates. External petroleum engineers may prepare more reliable reserve estimates if they constrain managers'

abilities to bias reserve estimates, but they may prepare less reliable reserve estimates if internal engineers possess private information they would otherwise use in estimating reserve quantities.

In addition to estimation error in reserve quantities, the SFAS No. 69 valuation model may also introduce measurement error into oil and gas asset present values. Adams et al. (1994) criticize the SFAS No. 69 valuation model because (1) it estimates future cash flows by multiplying the projected future production volumes times *current* oil and gas prices (i.e., the spot price on the balance sheet date) rather than *expected future* oil and gas prices, and (2) it discounts estimated future cash flows at a uniform rate of 10 percent rather than at a firm-specific discount rate. SFAS No. 69 effectively assumes that oil and gas prices do not change and that the appropriate discount rate neither varies across firms as a function of risk nor across time as a function of changes in market-wide interest rates. SFAS No. 69 thus mandates a present value model inconsistent with SFAC No. 7, which advocates explicitly recognizing uncertainty about the amount and timing of future cash flows by discounting the expected value of those cash flows at a firm-specific discount rate that reflects the risk associated with those cash flows as well as prevailing market rates of interest (FASB 2000, paras. 39, 40, 62–71).

These shortcomings of the SFAS No. 69 valuation model suggest that measurement error in the oil and gas present values varies across firm-years as a function of oil-price volatility and the deviation in firm-specific discount rates from 10 percent. Volatile spot prices increase the probability that prices will temporarily rest at an extreme value on the balance sheet date. If investors' expectations of future oil and gas prices are weakly related to spot prices on the balance sheet date, then the reported oil and gas present values are likely to differ from investors' implicit valuation of oil and gas assets, leading to greater measurement error in the present value estimates. In contrast, if investors' expectations of future oil and gas prices are strongly related to spot prices, then the measurement error in historical cost will be higher relative to the measurement error in present value because historical costs remain fixed while present values, a function of spot prices, move in synchrony with investors' implicit valuation of oil and gas assets. Therefore, oil and gas price volatility may lead to differential measurement error, but the direction is unclear *ex ante*. With respect to measurement error induced by the uniform 10 percent discount rate, the present value measure is more likely to deviate from the value investors assign oil and gas assets when the firm-specific discount rate differs from 10 percent. Based on the preceding discussion, I propose the following measurement error hypotheses (alternative form):

- H1:** Present value measurement error variance differs from historical cost measurement error variance in oil and gas assets.
- H1a:** Present value measurement error variance will be larger in firms that, on average, report larger annual revisions to the previous year's reserve quantity estimates than in firms that, on average, report smaller annual revisions.
- H1b:** Present value measurement error variance in firms employing independent (external) petroleum engineers to estimate reserve quantities will differ from present value measurement error variance in firms using internal petroleum engineers.
- H1c:** Present value measurement error variance when oil prices are volatile will differ from present value measurement error variance when oil prices are less volatile.³

³ I do not propose a similar hypothesis for gas price volatility because I was unable to obtain natural-gas spot prices for much of my sample.

H1d: Present value measurement error variance will be greater when firm-specific discount rates substantially differ from 10 percent than when close to 10 percent.

Model Misspecification Hypothesis

Magliolo (1986), Harris and Ohlson (1987), and Shaw and Wier (1993) rely on annual, cross-sectional estimates of an “imputed value” regression model that is a restricted form of the standard balance sheet valuation model. This model imputes the value of oil and gas assets as total market value of equity less the book value of non-oil-and-gas assets and liabilities (when liabilities are defined as negative values). The imputed value is then regressed on the accounting measure of oil and gas assets (either historical cost or present value). This specification restricts the coefficients on liabilities and non-oil-and-gas assets to 1 and the intercept to be constant across firms. The reportedly weak association between stock prices and present value (relative to historical cost) oil and gas asset measures may arise, in part, from model misspecification if the restrictions in the imputed value model are invalid—for example, when market value differs from book value of non-oil-and-gas assets and liabilities. I thus propose the following hypothesis:

H2: Specification error in the imputed value model contributes to the weak association between stock prices and oil and gas asset present values documented in prior research.

Time-Period Hypothesis

Prior studies (i.e., Magliolo 1986; Harris and Ohlson 1987; Doran et al. 1988; Alciatore 1993; Shaw and Wier 1993) are based on data from 1979–1988, suggesting that time-period idiosyncrasies may have been responsible for the reported weak value relevance of oil and gas present values. Thus, weak value relevance may not generalize to more recent years. This leads to my final hypothesis:

H3: The weak association between stock prices and oil and gas present values does not generalize to more recent years.

III. RESEARCH DESIGN

Test of H1

I assume a conventional cross-sectional valuation framework (Landsman 1986; Barth 1991, 1994) in which the market value of equity equals the difference between the summed market value of assets and liabilities, where market values are defined as the amounts investors implicitly assign to assets and liabilities when they value the firm:

$$MVE = OTHER\_MV + OG\_MV + LIAB\_MV \quad (1)$$

MVE is market value of equity, OTHER_MV is market value of all non-oil-and-gas assets, OG_MV is market value of oil and gas assets, and LIAB_MV is market value of liabilities (I consider liabilities negative values).⁴

⁴ I do not assume that market values implicit in share prices are unbiased or error-free measures of economic assets or liabilities. Rather, they represent benchmarks against which I assess measurement error.

Financial statements and related disclosures measure the market values of assets and liabilities with error:

$$\begin{aligned}\text{OTHER\_HC} &= \text{OTHER\_MV} + u_1 \\ \text{OG\_i} &= S_i \times (\text{OG\_MV} + u_{2i}) \\ \text{LIAB\_HC} &= \text{LIAB\_MV} + u_3\end{aligned}\tag{2}$$

where:

- OTHER_HC = the (historical cost) book value of non-oil-and-gas assets;
- OTHER_MV = the value that investors implicitly assign to non-oil-and-gas assets;
- OG_i = the accounting measure of oil and gas assets (either historical cost book value as denoted by $i = \text{HC}$, or the SFAS No. 69 present value measure as denoted by $i = \text{PV}$);
- OG_MV = the value that investors implicitly assign to oil and gas assets;
- LIAB_HC = the (historical cost) book value of liabilities multiplied by -1 ;
- LIAB_MV = the value that investors implicitly assign to liabilities multiplied by -1 ;
- u_1, u_{2i} , and u_3 = error in the accounting measure of the related market value construct; and
- S_i = a parameter reflecting the difference in scale between OG_HC and OG_PV.

Each error term, u_x , is a random variable with variance σ_x^2 , so the smaller the measurement error variance the more reliable the measure. Covariance between u_x and u_y is denoted $\sigma_{x,y}$. Equation (2) includes a scale factor S_i because OG_HC is systematically smaller than OG_PV. Empirically, OG_PV is 1.30 to 1.40 times larger than OG_HC (see Table 1, discussed later).

Measurement error, as defined by Equation (2), arises when accounting values and market values become misaligned. Historical costs and market values become misaligned because interest rate shocks or cash flow shocks after initial acquisition cause market values to change, while historical cost metrics remain largely fixed at initial acquisition price or at depreciated costs that approximate market value only by coincidence. The misalignment should be more severe in OG_HC because historical cost does not equal fair value, even at initial acquisition. However, because full cost accounting yields asset measures that are less severely understated than does successful efforts accounting, there should be less misalignment between OG_HC and OG_MV under full cost accounting.

Misalignment between market values and SFAS No. 69 present values also occurs, but the misalignment arises principally from interest rate effects (i.e., firm-specific discount rates that differ from 10 percent) rather than cash flow effects because petroleum firms update OG_PV at each balance sheet date to reflect changes in expected future cash flows. However, misalignment from cash flow shocks *can* arise if investors' expectations of future cash flows differ from the SFAS No. 69 expected amount, which might occur if, for example, investors suspect error in estimating reserve quantities, or if investors' expectations of future oil and gas prices differ from the spot prices at the balance sheet date SFAS No. 69 uses to value future oil and gas production. The Appendix provides greater detail about the nature of the measurement error terms.

Empirical estimation of Equation (1) with accounting values substituted for the unobservable market values yields the following errors-in-variables regression model:

$$MVE = \gamma_0 + \gamma_1 \text{OTHER\_HC} + \gamma_2 \text{OG\_i} + \gamma_3 \text{LIAB\_HC} + \text{HC} + \xi \quad (3)$$

where $\xi = -\gamma_1 u_1 - \gamma_2 S_i u_{2i} - \gamma_3 u_3$. Absent error in the accounting measures, the intercept in Equation (3) should be 0 and the slope coefficients on the regressors should be 1. With measurement error the expected value of the γ_k coefficients is not 1, but is $1 - B_k$ where B_k is the magnitude of bias in the coefficient. As the Appendix shows, the degree of bias B_k depends upon the variance/covariance structure in the accounting measurement error terms u_x .

I exploit this bias to structure a test of H1. Intuitively, my approach, based on Barth (1991) and Choi et al. (1997), involves first estimating the model in Equation (3) with OG_i defined as OG_PV and then again with OG_i defined as OG_HC. I then use the extent of observed bias in each estimation and a specification of the determinants of that bias to draw inferences about the relative magnitude of measurement error in the two alternative measures. As discussed in the Appendix, I implement this approach through a two-stage estimation technique. Coefficients estimated in the second stage form the basis for a statistical test of H1. The second stage model is:

$$\begin{aligned} MVE1^* &= \phi_j + \phi_t + \phi_{1,FC}(FC \times ZI_{\text{OTHER\_HC}}) + \phi_{2,FC}(FC \times ZI_{\text{OG\_PV}}) \\ &\quad + \phi_{3,FC}(FC \times ZI_{\text{LIAB\_HC}}) + \phi_{1,SE}(SE \times ZI_{\text{OTHER\_HC}}) \\ &\quad + \phi_{2,SE}(SE \times ZI_{\text{OG\_PV}}) + \phi_{3,SE}(SE \times ZI_{\text{LIAB\_HC}}) + \varepsilon_1 \\ MVE2^* &= \delta_j + \delta_t + \delta_{1,FC}(FC \times Z2_{\text{OTHER\_HC}}) + \delta_{2,FC}(FC \times Z2_{\text{OG\_HC}}) \\ &\quad + \delta_{3,FC}(FC \times Z2_{\text{LIAB\_HC}}) + \delta_{1,SE}(SE \times Z2_{\text{OTHER\_HC}}) \\ &\quad + \delta_{2,SE}(SE \times Z2_{\text{OG\_HC}}) + \delta_{3,SE}(SE \times Z2_{\text{LIAB\_HC}}) + \varepsilon_2 \end{aligned} \quad (4)$$

The parameters ϕ_j , ϕ_t , δ_j , and δ_t in Equation (4) denote firm (ϕ_j , δ_j) and year (ϕ_t , δ_t) fixed effects. FC and SE are 0,1 variables indicating the full cost or successful efforts method. The FC and SE variables allow the slope coefficients to differ between accounting methods because, as previously noted, the misalignment between OG_HC and OG_MV should be less severe under full cost than successful efforts accounting. MVE1* equals $MVE - (\text{OTHER\_HC} + \text{OG\_PV} + \text{LIAB\_HC})$ and MVE2* equals $MVE - ((S_{PV}/S_{HC}) \times \text{OG\_HC}) + \text{LIAB\_HC}$. The Z regressors are orthogonal values of OTHER_HC, $((S_{PV}/S_{HC}) \times \text{OG\_HC})$, OG_PV, and LIAB_HC obtained from the first-stage regressions.

S_{PV}/S_{HC} is a rescaling factor that controls for test bias due to a systematic scale difference between OG_HC and OG_PV. I calculate it by dividing the mean value of OG_PV by the mean value of OG_HC. Scale differences between OG_PV and OG_HC would confound comparisons of measurement error variance because variance is a function of scale (i.e., $\text{Var}(sx) = s^2 \text{Var}(x)$). Multiplying OG_HC by S_{PV}/S_{HC} transforms u_{2HC} to be of the same scale as u_{2PV} , and thus facilitates a scale-free comparison of measurement error variance between OG_HC and OG_PV (see the Appendix).⁵ Consequently, I ran all measurement error regressions after substituting $(S_{PV}/S_{HC}) \times \text{OG\_HC}$ in place of OG_HC.

⁵ I repeated the tests without the S_{PV}/S_{HC} rescaling factor. Results were qualitatively unchanged.

I estimate the model in Equation (4) as a seemingly unrelated regression, deflating each variable by barrel-of-oil equivalent units (BOE) to avoid size-related coefficient bias and mitigate heteroskedasticity.⁶ I pool observations across firms and years to increase statistical power, but I control for year-specific and firm-specific fixed effects. I also present annual cross-sectional regressions for comparison.

Subtracting δ_1 from ϕ_1 in Equation (4) and rearranging yields the following expression containing the two terms of interest, σ_{u2PV}^2 and σ_{u2HC}^2 :

$$\begin{aligned} \delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3 = & \sigma_{u2PV}^2 - \sigma_{u2HC}^2 + (\sigma_{\text{OTHER\_MV},u2HC} - \sigma_{\text{OTHER\_MV},u2PV} \\ & + \sigma_{\text{OG\_MV},u2PV} - \sigma_{\text{OG\_MV},u2HC} \\ & + \sigma_{\text{LIAB\_MV},u2HC} - \sigma_{\text{LIAB\_MV},u2PV}) \end{aligned} \quad (5)$$

A statistical test of H1, that $\sigma_{u2PV}^2 = \sigma_{u2HC}^2$, based on Equation (5), is $H_0: \delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3 = 0$. As detailed in the Appendix, the nature of historical cost in relation to present value asset measurement permits me to make plausible assumptions about the net signed value of the covariance terms in parentheses in Equation (5). In particular, I expect a positive, net value for the parenthetical covariance terms, and thus the coefficient function $\delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3$ is biased toward a positive value under the null hypothesis that measurement error variance in OG_PV and OG_HC are the same (i.e., $H_0: \sigma_{u2PV}^2 = \sigma_{u2HC}^2$). Consequently, a statistically significant *negative* value for the coefficient function in Equation (5) constitutes strong evidence to reject the null hypothesis that σ_{u2PV}^2 equals σ_{u2HC}^2 in favor of the alternative that σ_{u2PV}^2 is less than σ_{u2HC}^2 . However, a statistically significant *positive* value for the coefficient function provides weak evidence to reject the null hypothesis, because the positive value of the function could be attributable *either* to $\sigma_{u2PV}^2 > \sigma_{u2HC}^2$ *or* to the positive values of the covariance difference terms.

Tests of H1a–H1d

Hypotheses 1a–1d predict across-firm and across-time variation in present value measurement error based on hypothesized causal factors. I test these hypotheses by evaluating whether the coefficient function $\delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3$ differs between data subsets partitioned based on these factors.

I partition the data by adding to Equation (4) a dummy variable, D, both as a main effect and as an interaction term. D assumes a value of 1 or 0 depending upon the values of the partitioning variables EST_ERROR, INSIDE, OIL_VOLATILITY, and RATE, defined as follows.

EST_ERROR = proxy for the magnitude of the error in estimating reserve quantities, calculated as the absolute value of the coefficient of variation in a firm's revision to the beginning-of-year measure of OG_PV. I calculate this firm-specific measure using reserve revisions across the period 1984–1996. D = 1 if the value of EST_ERROR is above its pooled sample median value, and 0 otherwise;

⁶ Barrel-of-oil equivalent units (BOE), the deflator commonly used in prior oil and gas valuation studies, is a measure of the oil-equivalent energy content of oil and gas reserves. One calculates it by summing the physical units of oil (measured in barrels) and one-sixth of the physical units of natural gas (measured in thousands of cubic feet). (The thermal energy content of one thousand cubic feet of natural gas is approximately one-sixth the energy content of a barrel of oil.)

INSIDE = binary variable assigned a value of 1 if the firm relied on its own petroleum engineering staff to prepare reserve quantity estimates; 0 otherwise. D = 1 if INSIDE = 1, and 0 otherwise;

OIL_VOLATILITY = annualized standard deviation in the natural logarithm of the oil price relative, calculated as $\hat{\sigma}\sqrt{250}$, where $\hat{\sigma}$ is the estimated standard deviation of the log-transformed ratio of day t to day $t - 1$ spot price of oil. It is calculated across days 0 to -59 where day 0 corresponds to the balance sheet date. D = 1 if the value of OIL_VOLATILITY is above its pooled sample median value, and 0 otherwise; and

RATE = proxy for the degree of distortion in OG_PV introduced by the uniform 10 percent discount rate, calculated as $|r - 0.10|$, where r is the pension settlement rate the firm uses in accounting for defined benefit pension costs. D = 1 if the value of RATE is above its pooled sample median value, and 0 otherwise.

Test of H2

I base tests of H2 on Equations (6A) and (6B), which differ only in terms of the oil and gas measures (OG_PV vs. OG_HC). α_j and α_t in Equations (6A) and (6B) denote firm and time fixed effects. All other variables in Equations (6A) and (6B) have been previously defined.

$$\begin{aligned} \text{MVE} = & \alpha_j + \alpha_t + \gamma_{1t,FC}(\text{OTHER\_HC} \times \text{FC}) + \gamma_{2t,FC}(\text{OG\_PV} \times \text{FC}) \\ & + \gamma_{3t,FC}(\text{LIAB\_HC} \times \text{FC}) + \gamma_{1t,SE}(\text{OTHER\_HC} \times \text{SE}) \\ & + \gamma_{2t,SE}(\text{OG\_PV} \times \text{SE}) + \gamma_{3t,SE}(\text{LIAB\_HC} \times \text{SE}) + \varepsilon_1 \end{aligned} \quad (6A)$$

$$\begin{aligned} \text{MVE} = & \alpha_j + \alpha_t + \gamma_{1t,FC}(\text{OTHER\_HC} \times \text{FC}) + \gamma_{2t,FC}(\text{OG\_HC} \times \text{FC}) \\ & + \gamma_{3t,FC}(\text{LIAB\_HC} \times \text{FC}) + \gamma_{1t,SE}(\text{OTHER\_HC} \times \text{SE}) \\ & + \gamma_{2t,SE}(\text{OG\_HC} \times \text{SE}) + \gamma_{3t,SE}(\text{LIAB\_HC} \times \text{SE}) + \varepsilon_2 \end{aligned} \quad (6B)$$

Prior research estimated both Equations (6A) and (6B) as annual cross-sectional regressions, subject to the restriction that $\gamma_{1t} = \gamma_{3t} = 1$. Researchers also restricted the OG_PV coefficient to be the same for full cost and successful efforts firms—that is, they restricted $\gamma_{2t,FC} = \gamma_{2t,SE}$ in models corresponding to Equation (6A).

As a test of H2, I estimate Equations (6A) and (6B) and evaluate both the appropriateness of these coefficient restrictions and their impact on the explanatory power of the models. To facilitate comparison with prior research, I base the analysis on data for the 1984–88 subperiod, which are the years common to both my study and prior studies.

Test of H3

I test H3 by comparing the results obtained for the 1984–1996 full period to those for the 1984–88 subperiod that overlaps the period used in prior research and the 1989–1996 subperiod subsequent to prior research.

IV. SAMPLE SELECTION AND DATA

I merged the 1997 Compustat PC Plus database, the 1997 CRSP database, and the Arthur Andersen Oil and Gas Disclosures Database for the years 1988 (containing data for

1984–88), 1993 (containing data for 1989–1993), and 1996 (containing data for the years 1994–96), and retained all firm-years with nonmissing data.⁷ The final sample includes 745 firm-year observations from 111, SIC 1311 exploration and production firms. I exclude integrated/diversified petroleum firms in SIC 2911 to ensure a homogenous group of firms for which oil and gas assets comprise a substantial portion of firm value.

Table 1 presents descriptive statistics for all regression variables, expressed in dollars per barrel-of-oil equivalent units (BOE). The table also presents supplementary statistics on total market value of equity (SIZE), systematic risk (BETA), and the deflator (BOE). These results are consistent with prior research (Malmquist 1990; Deakin 1979) in that successful efforts firms (1) are larger than full cost firms (SIZE and BOE are larger in successful efforts firms) and (2) report lower capitalized oil and gas asset costs (OG_HC is smaller in successful efforts firms). Successful efforts firms also own less-valuable reserves (OG_PV is smaller in successful efforts firms) and rely less on debt financing than do full cost firms (absolute value of mean LIAB_HC is smaller in successful efforts firms).

Oil and gas assets comprise a large fraction of total assets (i.e., $OG\_PV/[OTHER\_HC + OG\_PV]$ is 70 percent in the successful efforts group and 72 percent in the full cost group). Thus, oil and gas assets are large enough to affect cross-sectional variation and stock prices. Moreover, OG_PV is systematically larger than OG_HC in both groups. The mean difference (OG_PV – OG_HC) is \$0.52 per barrel-of-oil-equivalent unit in the successful efforts group and \$0.42 per barrel-of-oil-equivalent unit in the full cost group. The mean ratio OG_PV/OG_HC exceeds 1 in both the successful efforts group (1.39) and in the full cost group (1.31), reflecting the scale-related differences between OG_HC and OG_PV.

V. EMPIRICAL RESULTS

Test of H1 Based on Measurement Error Model

Table 2 reports tests of H1 based on direct estimates of measurement error variance. Panel A reports the details of the parameter estimates of the measurement error model. Of greater interest, Panel B reports the key tests of the coefficient function $\delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3$. For full cost firms and successful efforts firms, the values of the coefficient functions are –1.52 and –0.83, respectively, both significantly negative ($Z = -12.01$ and $Z = -6.16$). The significantly negative coefficient function suggests that present value measures of oil and gas assets suffer from *less* measurement error variance than do historical cost measures (i.e., $\sigma_{u2PV}^2 < \sigma_{u2HC}^2$). This evidence is particularly compelling because, as already noted, the value of the coefficient function is biased toward a *positive* value.

The value of the coefficient function is significantly more negative for the full cost firms (–1.52) than for the successful efforts firms (–0.83) ($p < 0.001$ in an untabulated test). This result suggests that historical cost is less reliable relative to present value for the full cost firms than for the successful efforts firms. This result is surprising because, as noted above, successful efforts understates oil and gas asset historical costs more severely than does the full cost method, which suggests that the successful efforts method should suffer from more historical cost measurement error. One possible explanation is that successful efforts firms also have large present value measurement error variance, which may obscure the conjectured large historical cost measurement error variance. Table 1 reports

⁷ To retain a reasonable sample size, I do not exclude observations with missing values for the pension discount settlement rate. Thus, analysis of RATE partitions is based on the subset of 238 firm-year observations with available discount settlement rate data.

TABLE 1
Sample Descriptive Statistics and Tests for Differences in Means across Successful Efforts and Full Cost Firms^{a,b}

	Successful Efforts Firms				Full Cost Firms				Significance Level from Two-Tailed Test for Difference in Means between Successful Efforts and Full Cost Firms
	<i>n</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Median</i>	<i>n</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Median</i>	
MVE	370	4.48	2.52	4.06	375	5.31	4.05	4.13	0.00
OTHER_HC	370	2.11	2.17	1.27	375	2.28	2.79	1.43	0.36
OG_HC	370	3.47	1.27	3.43	375	4.15	1.77	4.10	0.00
OG_PV	370	3.98	3.99	3.78	375	4.57	1.68	4.30	0.00
LIAB_HC	370	-2.91	2.12	-2.61	375	-3.27	2.55	-2.64	0.04
OG_PV									
OTHER_HC + OG_PV	370	0.70	0.18	0.73	375	0.72	0.17	0.75	0.11
SIZE	370	836.41	1,723.43	136.47	375	217.86	454.19	57.84	0.00
BETA	370	0.76	0.70	0.78	375	0.79	0.82	0.76	0.51
BOE	370	191.40	384.14	29.25	375	42.03	78.82	14.99	0.00
OG_PV - OG_HC	370	0.52	1.42	0.40	375	0.42	1.67	0.25	0.42
OG_PV/OG_HC	370	1.39	1.32	1.12	375	1.31	0.91	1.07	0.31
EST_ERROR	370	5.35	129.65	2.45	375	4.16	52.27	2.39	0.08
INSIDE	370	0.21	0.41	0.00	375	0.11	0.31	0.00	0.00
OIL_VOLATILITY	370	13.99	11.36	12.75	375	16.21	15.27	12.75	0.03
RATE	132	0.04	0.03	0.03	106	0.07	0.04	0.10	0.00

(Continued on next page)

TABLE 1 (Continued)

^a n denotes number of firm-year observations.

^b Variable definitions:

MVE	= price per share of common stock (measured as of the first trading day 90 days after fiscal year-end) multiplied by numbers of shares outstanding, deflated by ending oil and gas reserves measured in barrel-of-oil equivalent units;
OTHER_HC	= the amortized historical cost (book value) of assets excluding oil and gas properties, deflated by ending oil and gas reserves measured in barrel-of-oil equivalent units;
OG_HC	= the amortized historical cost (book value) measure of oil and gas assets, deflated by ending oil and gas reserves measured in barrel-of-oil equivalent units;
OG_PV	= the SFAS No. 69 discounted present value of oil and gas assets, deflated by ending oil and gas reserves measured in barrel-of-oil equivalent units;
LIAB_HC	= the amortized historical cost (book value) of liabilities multiplied by -1, deflated by ending oil and gas reserves measured in barrel-of-oil equivalent units;
SIZE	= price per share of common stock (measured as of the first trading day 90 days after fiscal year-end) multiplied by shares outstanding (in millions);
BETA	= systematic risk estimated using weekly returns across the one-year period before the balance sheet date;
BOE	= barrel-of-oil equivalent units (in millions), calculated by summing the physical units of oil (measured in barrels) and one-sixth of the physical units of natural gas (measured in thousands of cubic feet);
EST_ERROR	= proxy for the magnitude of the error in estimating reserve quantities, and is the absolute value of coefficient of variation in a firm's revision to the beginning of year measure of OG_PV. It is a firm-specific measure calculated using the firm's reserve revisions across the period 1984-1996;
INSIDE	= a binary variable assigned a value of 1 if the firm relied on its own petroleum engineering staff to prepare reserve quantity estimates, and 0 otherwise;
OIL_VOLATILITY	= the annualized standard deviation in the natural logarithm of the oil price relative. It is calculated as $\hat{\sigma}\sqrt{250}$, where $\hat{\sigma}$ is the estimated standard deviation of the natural logarithm of the ratio of the spot price of oil on day t to the spot price of oil on day t - 1 calculated across days 0 to -59, where day 0 corresponds to the balance sheet date; and
RATE	= a proxy for the degree of distortion in OG_PV introduced by the uniform 10 percent discount rate, and is calculated as $ r - 0.10 $, where r is the pension settlement rate the firm uses in accounting for defined benefit pension costs.

TABLE 2
Pooled Fixed-Effects and Annual Regression Estimates of Measurement Error Model and Related Tests of the Reliability of the Present Value Measure of Oil and Gas Assets as Compared to the Reliability of the Historical Cost Measure of Oil and Gas Assets: 1984-1996

$$\begin{aligned} MVE1^* &= \phi_j + \phi_i + \phi_{1,FC}(FC \times Z1_{OTHER\_HC}) + \phi_{2,FC}(FC \times Z1_{OG\_PV}) + \phi_{3,FC}(FC \times Z1_{LIAB\_HC}) \\ &\quad + \phi_{1,SE}(SE \times Z1_{OTHER\_HC}) + \phi_{2,SE}(SE \times Z1_{OG\_PV}) + \phi_{3,SE}(SE \times Z1_{LIAB\_HC}) + \varepsilon_1 \\ MVE2^* &= \delta_j + \delta_i + \delta_{1,FC}(FC \times Z2_{OTHER\_HC}) + \delta_{2,FC}(FC \times Z2_{OG\_HC}) + \delta_{3,FC}(FC \times Z2_{LIAB\_HC}) \\ &\quad + \delta_{1,SE}(SE \times Z2_{OTHER\_HC}) + \delta_{2,SE}(SE \times Z2_{OG\_HC}) + \delta_{3,SE}(SE \times Z2_{LIAB\_HC}) + \varepsilon_2^a \end{aligned}$$

Panel A: Estimates of Measurement Error Model

		Pooled Two-Way Fixed- Effects Estimation	Annual Cross-Sectional Estimation				
Variable	Coefficient		Mean of Annual Estimates	Zl ^b	Z2 ^b	Significantly Different from 0 at ≤ 0.10 Level	
						Num. > 0	Num. < 0
FC × Zl _{OTHERASSETS_HC}	φ _{1FC}	2.11 (4.34)***	-3.40 (-0.94)	-3.37***	-1.49	2	5
FC × Zl _{O&GASSETS_PV}	φ _{2FC}	2.80 (23.61)***	1.87 (7.35)	26.50***	5.30***	11	0
FC × Zl _{LIAB_HC}	φ _{3FC}	-1.30 (-2.84)***	3.55 (1.60)	5.76***	3.23***	7	0
SE × Zl _{OTHERASSETS_HC}	φ _{1SE}	-0.87 (-1.67)*	-0.56 (-0.17)	-0.60	-0.40	1	1
SE × Zl _{O&GASSETS_PV}	φ _{2SE}	1.64 (26.93)***	1.63 (41.09)	148.15***	14.17***	13	0

(Continued on next page)

TABLE 2 (Continued)

Variable	Coefficient	Pooled Two-Way Fixed- Effects Estimation	Annual Cross-Sectional Estimation			
			Mean of Annual Estimates	Z1 ^b	Z2 ^b	Significantly Different from 0 at ≤ 0.10 Level
SE × Z1 _{LIAB_HC}	ϕ _{3,SE}	1.01 (1.90)*	1.83 (1.54)	5.55***	6.76***	Num. > 0 5 Num. < 0 0
FC × Z2 _{OTHERASSETS_HC}	δ _{1,FC}	3.25 (6.57)***	-2.26 (-0.21)	-0.74	-0.29	3 5
FC × Z2 _{O&GASSETS_HC}	δ _{2,FC}	3.76 (16.73)***	1.66 (3.25)	11.73***	3.87***	9 1
FC × Z2 _{LIAB_HC}	δ _{3,FC}	-2.99 (-6.44)***	1.86 (0.47)	1.68*	0.84	3 3
SE × Z2 _{OTHERASSETS_HC}	δ _{1,SE}	-0.49 (-0.92)	-0.18 (0.17)	0.61	0.39	2 1
SE × Z2 _{O&GASSETS_HC}	δ _{2,SE}	1.85 (11.52)***	1.52 (5.20)	18.75***	10.74***	13 0
SE × Z2 _{LIAB_HC}	δ _{3,SE}	0.01 (0.02)	0.83 (0.61)	2.20**	2.55***	3 0

(Continued on next page)

TABLE 2 (Continued)

Panel B: Test of H1, Specified in Null Form as $\delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3 = 0$		Annual Cross-Sectional Estimation			
Description	Pooled Two-Way Fixed-Effects Estimation	Mean of Annual Estimates	Significantly Different from 0 at ≤ 0.10 Level		
			Z1*	Z2*	Num. > 0 Num. < 0
Value of $\delta_{1,FC} - \delta_{2,FC} + \delta_{3,FC} - \phi_{1,FC} + \phi_{2,FC} - \phi_{3,FC}$	-1.52	-0.34			
Z-statistic testing the null hypothesis in FC firms	-12.01***	-1.48	-5.33***	-2.32***	2 7
Value of $\delta_{1,SE} - \delta_{2,SE} + \delta_{3,SE} - \phi_{1,SE} + \phi_{2,SE} - \phi_{3,SE}$	-0.83	-0.51			
Z-statistic testing the null hypothesis in SE firms	-6.16***	-2.03	-7.32***	-6.41***	0 9

***, **, * Denote significance in two-tailed tests at the 0.10, 0.05, and 0.01 levels, respectively.

^aIn the fixed-effects estimation, the parameters ϕ_j and δ_j denote firm fixed effects and ϕ_j and δ_j are year fixed effects. In the annual cross-sectional estimation, ϕ_j and δ_j denote separate intercepts for firms in the full cost and successful efforts groups. For brevity, the estimated values of these parameters are not reported.

FC and SE are dummy variables assigned a value of 1 if the firm uses the full cost or successful efforts method, respectively, and 0 otherwise.

The Z regressors are orthogonal values of OTHER_HC, OG_PV, OG_HC, and LIAB_HC obtained from first-stage regressions as shown in the Appendix.

All variables are deflated by barrel-of-oil equivalent units.

MVE1* equals $MVE - (OTHER\_HC + OG\_PV + LIAB\_HC)$ and MVE2* equals $MVE - (OTHER\_HC + (S_{PV}/S_{HC}) \times OG\_HC) + LIAB\_HC$.

^bAll other variables are defined in Table 1.

Z1 and Z2 statistics test whether the time-series mean asymptotic t-statistic equals 0. Z1 equals $1/\sqrt{N} \sum_{j=1}^N t_j / \sqrt{k_j / (k_j - 2)}$ where t_j is the asymptotic t-statistic for year j , k_j is the degrees of freedom for year j , and N is the number of years. Z2 equals $\bar{t} / (\text{stddev}(t) / \sqrt{N - 1})$.

^cZ1* and Z2* statistics test whether the time-series mean Z-statistic equals 0. Z1* equals $\bar{z} / \sqrt{N}$ and Z2* equals $\bar{z} / (\text{stddev}(z) / \sqrt{N - 1})$. The tests reported in Panel B are based on a χ^2 statistic with 1 degree of freedom obtained from a Wald test. I convert the χ^2 statistic to an equivalent Z-statistic ($z = \sqrt{\chi^2 / (1 \text{ df})}$) to construct the Z1* and Z2* statistics.

evidence consistent with this speculation, as successful efforts firms have significantly greater reserve quantity estimation error (EST_ERROR) than do full cost firms (5.35 vs. 4.16, significant at the 0.08 level). Tests of H1a, discussed in the following subsection, document that firms with more reserve quantity estimation error suffer from larger measurement error in their present values.

In annual cross-sectional estimates, the mean value of the coefficient function is -0.34 and -0.51 for full cost and successful efforts firms, respectively (third column in Table 2, Panel B). The coefficient function is significantly negative in seven of the 13 annual tests for the full cost firms and in nine of the 13 annual tests for the successful efforts firms (last column of Panel B). The mean Z-statistics from these annual tests ($\bar{Z} = -1.48$ and -2.03 in the third column of Panel B) are significantly negative (tests based on Z1* and Z2* statistics in Panel B).⁸ Consequently, these annual cross-sectional estimates also indicate that present value oil and gas measures suffer from *less* measurement error than do historical cost measures.

In summary, the evidence in Table 2, from both pooled, fixed-effects regressions and from annual regressions, consistently supports H1. The results indicate that historical cost oil and gas asset measures are noisier (less reliable) than present value measures.

Tests of H1a–H1d Based on Measurement Error Model

Table 3 summarizes tests of H1a–H1d based on the measurement error model, augmented with the variable D added both as a main effect and as an interaction term. For brevity, I suppress the individual coefficient estimates. A test that the coefficient function $\delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3$ differs between partitions is $H_0: \delta_4 - \delta_5 + \delta_6 - \phi_4 + \phi_5 - \phi_6 = 0$. A negative value for $\delta_4 - \delta_5 + \delta_6 - \phi_4 + \phi_5 - \phi_6$ implies less present-value measurement error variance in the $D = 1$ group relative to the $D = 0$ group, and a positive value implies the opposite.

Panel A summarizes test results for full cost firms and Panel B summarizes the results for successful efforts firms. In Panel A, the significantly positive value of the coefficient function $\delta_4 - \delta_5 + \delta_6 - \phi_4 + \phi_5 - \phi_6$ in the “EST_ERROR” column indicates that present value measures suffer from more measurement error in firms that routinely report larger annual revisions to the previous year’s reserve quantity estimates. Similarly, the significantly positive value of the coefficient function in the “RATE” column indicates that present value measures suffer from more measurement error when firm-specific discount rates differ substantially from the required, uniform 10 percent discount rate than when rates are close to 10 percent. The coefficient function is not significantly different from 0 for full cost firms in the “INSIDE” column (test of H1b) or the “OIL_VOLATILITY” column (test of H1c). Thus, there is no evidence that present value measurement error variance is related to use of external petroleum engineers or to oil price volatility. Finally, the coefficient function is not significantly different from 0 in successful efforts firms for any of the partitioning variables (Panel B).

In summary, although there is no evidence that present value measures of oil and gas assets are significantly noisier when internal engineers prepare the reserve estimates or when oil prices are more volatile, evidence does indicate that present value measures are noisier in full cost firms (1) experiencing large reserve estimate revisions and (2) when firm-specific discount rates are substantially different from the SFAS No. 69 uniform 10 percent discount rate. Thus, tests support H1a and H1d, but only in full cost firms.

⁸ Z1* equals $\bar{z}\sqrt{N}$ and Z2* equals $\bar{z}/(\text{stddev}(z)/\sqrt{(N-1)})$. Z1* assumes independence in the annual Z-statistics, whereas Z2* corrects for potential lack of independence (Healy et al. 1987; Christie 1990; Barth 1994).

TABLE 3
Pooled Fixed-Effects Test for Differences in the Reliability of Oil and Gas Assets Measured at Present Value as Compared to the Reliability When Measured at Historical Cost—For Data Partitioned Based on Factors Expected to Influence the Across-Firm and Across-Time Reliability of the Present Value Measure: 1984–1996^a

$$\begin{aligned} MVE1^* &= \phi_j + \phi_i + \phi_0 D + \phi_{1,FC}(FC \times Z1_{OTHER\_HC}) + \phi_{2,FC}(FC \times Z1_{OG\_PV}) \\ &\quad + \phi_{3,FC}(FC \times Z1_{LIAB\_HC}) + \phi_{1,SE}(SE \times Z1_{OTHER\_HC}) + \phi_{2,SE}(SE \times Z1_{OG\_PV}) \\ &\quad + \phi_{3,SE}(SE \times Z1_{LIAB\_HC}) + \phi_{4,FC}(D \times FC \times Z1_{OTHER\_HC}) + \phi_{5,FC}(D \times FC \times Z1_{OG\_PV}) \\ &\quad + \phi_{6,FC}(D \times FC \times Z1_{LIAB\_HC}) + \phi_{4,SE}(D \times SE \times Z1_{OTHER\_HC}) + \phi_{5,SE}(D \times SE \times Z1_{OG\_PV}) \\ &\quad + \phi_{6,SE}(D \times SE \times Z1_{LIAB\_HC}) + \epsilon_1 \\ MVE2^* &= \delta_j + \delta_i + \delta_0 D + \delta_{1,FC}(FC \times Z2_{OTHER\_HC}) + \delta_{2,FC}(FC \times Z2_{OG\_HC}) \\ &\quad + \delta_{3,FC}(FC \times Z2_{LIAB\_HC}) + \delta_{1,SE}(SE \times Z2_{OTHER\_HC}) + \delta_{2,SE}(SE \times Z2_{OG\_HC}) + \delta_{3,SE}(SE \times Z2_{LIAB\_HC}) \\ &\quad + \delta_{4,FC}(D \times FC \times Z2_{OTHER\_HC}) + \delta_{5,FC}(D \times FC \times Z2_{OG\_HC}) + \delta_{6,FC}(D \times FC \times Z2_{LIAB\_HC}) \\ &\quad + \delta_{4,SE}(D \times SE \times Z2_{OTHER\_HC}) + \delta_{5,SE}(D \times SE \times Z2_{OG\_HC}) + \delta_{6,SE}(D \times SE \times Z2_{LIAB\_HC}) + \epsilon_2 \end{aligned}$$

Panel A: Tests for Full Cost Firms

Partitioning Variable		
EST_ERROR (Test of H1a)	INSIDE (Test of H1b)	OIL_VOLATILITY (Test of H1c)
RATE (Test of H1d)		

Test for a difference between $D = 1$ and $D = 0$ in reliability of oil and gas assets measured at present value as compared to reliability when measured at historical cost. Test specified as:

$$\delta_{4,FC} - \delta_{3,FC} + \delta_{6,FC} - \phi_{4,FC} + \phi_{5,FC} - \phi_{6,FC} = 0$$

Value of coefficient function
 χ^2 statistic from a Wald test of null hypothesis

0.89
3.34**

(Continued on next page)

Tests of H2

Table 4 presents specification tests of Equations (6A) (present value model) and (6B) (historical cost model) for the years 1984–88, which are the years that overlap prior research sample periods. Consistent with prior research, the “Restricted Model” restricts the slope coefficients on OG_PV to be the same ($\gamma_{2tFC} = \gamma_{2tSE}$) in Equation (6A) only, and restricts the slope coefficients on LIAB_HC and OTHER_HC to a value of 1 in both Equations (6A) and (6B) ($\gamma_{1t} = \gamma_{3t} = 1$) (Magliolo 1986; Harris and Ohlson 1987; Shaw and Wier 1993).

The most general specification, Model 1, provides for fixed time and firm effects and allows slope coefficients $\gamma_{k,t}$ to vary across years. In unrestricted estimates of Model 1, the present value specification's R^2 significantly exceeds the historical cost specification's R^2 ($0.83 > 0.81$, $p = 0.07$). In restricted estimates of Model 1, the R^2 is not significantly different between the present value and historical cost models. The right-hand columns of Table 4 show that the restrictions are inappropriate and that the unrestricted Model 1 has significantly greater explanatory power than the restricted Model 1 in both the present value and historical cost models ($p \leq 0.001$). Results are similar in Model 2 that holds coefficients constant across years.

Results differ markedly, however, in specifications without firm fixed effects, Models 3–8, that correspond to specifications estimated in prior research. Model 3 differs from Models 4–8 only in that Model 3 constrains the variance of the disturbance terms to be temporally constant, whereas the variance is free to vary across years in Models 4–8. In unrestricted estimates of Model 3, the R^2 is not significantly different between historical cost and present value specifications. In restricted estimates of Model 3 (the specification used in prior studies), the historical cost specification's R^2 significantly exceeds the present value specification's R^2 ($p < 0.05$). This finding replicates results reported in Harris and Ohlson (1987) and Shaw and Wier (1993). Importantly, however, the right-hand columns of Table 4 show that the restrictions are inappropriate and that the unrestricted version of Model 3 has significantly greater explanatory power than the restricted version in both present value and historical cost specifications ($p < 0.001$).

In annual unrestricted estimates of Models 4–8, the differences in R^2 values are not significant in Vuong (1989) tests. In annual restricted estimates of Models 4–8, the present value specification's R^2 is significantly *less* than the historical cost specification's R^2 in 1986 and 1987, and marginally so in 1988. As in Model 3, these findings confirm prior research. Also as in Model 3, however, tests of the restricted vs. unrestricted specifications of Models 4–8 (reported in the right-hand columns of Table 4) reveal that the restrictions are inappropriate in 1986, 1987, and 1988 in the historical cost model, and in 1985–88 in the present value model.

A test of Model 1 vs. Model 2 (bottom of Table 4) cannot reject the hypothesis of temporal stability in slope coefficients. In contrast, a test of Model 1 vs. Model 3 rejects the hypothesis of no firm fixed effects in all specifications at the 0.001 level. Thus, specification tests indicate that unrestricted Models 1 and 2 are better specified than the competing Models 3–8 that do not allow firm fixed effects.

In summary, restricted versions of Models 3, 5, 6, 7, and 8 that correspond to the models used in prior research replicate prior research's reportedly weak association between stock prices and present value (relative to historical cost) oil and gas asset measures. However, specification tests reject these misspecified models in favor of unrestricted forms of Model 1 or Model 2 that: (1) allow for firm fixed effects and (2) do not constrain the slope coefficients on LIAB_HC and OTHER_HC to equal 1, nor do they constrain OG_PV to be the same for full cost and successful efforts firms. In the more appropriately specified

TABLE 4
Tests of Model Specification, 1984–88

Present Value Model: $MVE = \alpha_j + \alpha_i + \gamma_{11,FC}(OTHER\_HC \times FC) + \gamma_{21,FC}(OG\_PV \times FC) + \gamma_{31,FC}(LIAB\_HC \times FC) + \gamma_{11,SE}(OTHER\_HC \times SE) + \gamma_{21,SE}(OG\_PV \times SE) + \gamma_{31,SE}(LIAB\_HC \times SE) + \varepsilon_1$

Historical Cost Model: $MVE = \alpha_j + \alpha_i + \gamma_{11,FC}(OTHER\_HC \times FC) + \gamma_{21,FC}(OG\_HC \times FC) + \gamma_{31,FC}(LIAB\_HC \times FC) + \gamma_{11,SE}(OTHER\_HC \times SE) + \gamma_{21,SE}(OG\_HC \times SE) + \gamma_{31,SE}(LIAB\_HC \times SE) + \varepsilon_2$

Significance Level from Test of Restricted vs. Unrestricted Model^b

Model	Description	Unrestricted Model ^a			Restricted Model ^a			Unrestricted Model ^b		
		R ² Present Value Model	R ² Historical Cost Model	Significance Levels from Vuong Test of $R^2_{PV} = R^2_{HC}$	R ² Present Value Model	R ² Historical Cost Model	Significance Levels from Vuong Test of $R^2_{PV} = R^2_{HC}$	Present Value Model	Historical Cost Model	Model
1	Pooled with firm and time fixed effects and time-varying slopes ($\alpha_j \neq \alpha_i$; $\alpha_i \neq \alpha_{t+1}$; $\gamma_{k,t} \neq \gamma_{k,t+1}$)	0.83	0.81	0.07	0.74	0.73	0.97	0.00	0.00	0.00
2	Pooled with firm and time fixed effects but slopes held constant across time ($\alpha_j \neq \alpha_i$; $\alpha_i \neq \alpha_{t+1}$; $\gamma_{k,t} = \gamma_{k,t+1}$)	0.77	0.74	0.11	0.74	0.71	0.18	0.01	0.08	0.08
3	Pooled with time-varying slopes and time fixed effects but no firm fixed effects ($\alpha_j = \alpha_i$; $\alpha_i \neq \alpha_{t+1}$; $\gamma_{k,t} \neq \gamma_{k,t+1}$)	0.45	0.42	0.58	0.17	0.28	0.05	0.00	0.00	0.00

(Continued on next page)

TABLE 4 (Continued)

Model	Description	Unrestricted Model ^a				Restricted Model ^a				Significance Level from Test of Restricted vs. Unrestricted Model ^b	
		R ² Present Value Model	R ² Historical Cost Model	Significance Levels from Vuong Test of $R^2_{PV} = R^2_{HC}$	R ² Present Value Model	R ² Historical Cost Model	Significance Levels from Vuong Test of $R^2_{PV} = R^2_{HC}$	Present Value Model	Hist. Cost Model		
4	1984 cross-sectional	0.49	0.29	0.50	0.53	0.34	0.42	0.55	0.51		
5	1985 cross-sectional	0.81	0.77	0.55	0.54	0.62	0.52	0.02	0.20		
6	1986 cross-sectional	0.36	0.18	0.32	0.02	0.19	0.07	0.00	0.09		
7	1987 cross-sectional	0.55	0.50	0.39	0.02	0.21	0.10	0.00	0.01		
8	1988 cross-sectional	0.25	0.34	0.26	0.02	0.22	0.12	0.03	0.08		
		Significance Levels from Model Comparisons			Significance Levels from Model Comparisons						
Model Comparisons		PV	HC		PV	HC					
1 vs. 2		0.22	0.67		0.21	0.13					
1 vs. 3		0.00	0.00		0.00	0.00					

^a The parameters α_j and α denote firm and year fixed effects; γ_{ki} denotes slope coefficients free to vary by year. Consistent with prior research, the restricted model requires $\gamma_{1i} = \gamma_{2i} = \gamma_{3i} = 1$ in both the historical cost and present value models and $\gamma_{2i,HC} = \gamma_{2i,SE}$ in the present value model.

^b Test based on following F-statistic: $F[J, n - K] = [(SSE_R - SSE_{UR})/J]/[SSE_{UR}/(n - K)]$. SSE denotes sum of squared residuals for restricted (R) and unrestricted (UR) models. J denotes the number of restrictions, K is the number of parameters, and n is the number of observations.

^a The parameters α_i and α_i denote firm and year fixed effects; γ_{it} denotes slope coefficients free to vary by year. Consistent with prior research, the restricted model requires $\gamma_{it} = \gamma_{it} = 1$ in both the historical cost and present value models and $\gamma_{it,FC} = \gamma_{it,SE}$ in the present value model.

^b Test based on following F-statistic: $F[J, n - K] = [(SSE_R - SSE_{UR})/J]/[SSE_{UR}/(n - K)]$. SSE denotes sum of squared residuals for restricted (R) and unrestricted (UR) models, J denotes the number of restrictions, K is the number of parameters, and n is the number of observations.

Models 1 and 2, present value measures of oil and gas assets have significantly greater explanatory power for stock prices than do historical cost measures. Thus, the weak association between stock prices and present value (relative to historical cost) oil and gas asset measures during the period 1984–88 seems to be attributable to model misspecification that disappears when the specification error is corrected. In brief, evidence obtained from model specification tests supports H2.

Why do the coefficient restrictions imposed in annual, cross-sectional imputed value models so severely diminish the association between stock prices and the present value measures of oil and gas assets, and why do pooled, fixed-effect regressions attenuate this effect? Although I leave it to future research to answer this question, I offer two conjectures. First, measurement error in the present value measure may be correlated with measurement error in the non-oil-and-gas assets and liabilities, causing the slope coefficients on non-oil-and-gas assets and liabilities to deviate from their theoretical expected value of 1. In this case, the effect of measurement error in non-oil-and-gas assets and liabilities, although by itself perhaps weak, may be exacerbated by covariance with present value measurement error. The weak results from the restricted annual valuation estimates may thus be attributable to a defect in the present value measure that is offset either by relaxing the coefficient restriction or by estimating a fixed-effect specification. Alternatively, if error in the present value measure is independent of error in the measure of non-oil-and-gas assets and liabilities, then error in measuring non-oil-and-gas assets and liabilities alone accounts for the weak explanatory power of present value in the restricted annual estimates. In this case, relaxing the coefficient restrictions or estimating a fixed-effect specification compensates, in effect, for defects in the measurement of non-oil-and-gas assets and liabilities.

Test of H3

Hypothesis 3 proposes that the weak value relevance of oil and gas present values as documented in prior research does not generalize to more recent years. The tests of H2 reported in Table 4 reveal that the weak value relevance of oil and gas present values in 1984–88 disappears after I correct apparent model misspecification. Nonetheless, for the sake of completeness, I estimate (in unrestricted form) Equations (6A) and (6B) using data for the 1984–1996 full period, the 1984–88 subperiod, and the 1989–1996 subperiod. In each period, the present value specification's R^2 significantly exceeds the historical cost specification's R^2 ($p < 0.10$ in untabulated tests). Thus, results that hold for the full period 1984–1996 also hold for the 1984–88 and the 1989–1996 subperiods, so the data do not support H3.

Additional Analysis

Because the measurement error analysis requires strong and untestable assumptions, I provide corroborative evidence from alternative supplementary tests of measurement error based on R^2 comparisons. Assuming measurement error reduces R^2 in addition to causing coefficient bias, I interpret model specifications with the smaller R^2 as suffering from more measurement error. I compare the R^2 from Equation (6A) (present value model) to the R^2 from Equation (6B) (historical cost model) using Vuong's (1989) likelihood ratio test. This supplementary analysis of R^2 differs from the pooled analysis of R^2 reported in Table 4 in that: (1) the period examined is longer (1984–1996 vs. 1984–1988 in Table 4) and (2) the data are partitioned by factors expected to influence the across-firm and across-time reliability of the present value measure.

One should interpret the R^2 comparisons cautiously, however, because R^2 values do not perfectly rank order alternative accounting metrics according to relative measurement-error

variance unless one of the compared models is measured without error and the measurement error is independent of the true value (Barth 1989). This assumption is *not* strictly met here because I assume that all accounting metrics are noisy measures of market values, and I do not assume independence between measurement error and market values. Thus, these results are only suggestive.

Test of H1 Based on R² Comparisons

Table 5 presents the supplementary tests of H1 based on R² comparisons. In the unrestricted fixed-effects model (Model 2 from Table 4), the present value specification's R² (0.68) is significantly higher than the historical cost specification's R² (0.66) per the Vuong (1989) tests ($p = 0.08$ level). This additional evidence that the present value specification suffers from less measurement error than the historical cost specification confirms the main tests of H1.

Similarly, in annual cross-sectional estimates, the mean R² from the present value specification is 0.36 and the mean R² from the historical cost specification is 0.32. Vuong tests based on annual estimates reject $H_0: R^2_{PV} = R^2_{HC}$ in 3 out of the 13 years (untabulated). The mean Z-statistic obtained from these annual Vuong tests is $\bar{Z} = 0.53$ (untabulated), which is significant in a test based on Z1* ($p = 0.06$), but not in a test based on Z2* ($p = 0.12$).⁹ Because Z2* is a more robust test than Z1* (albeit a weaker test), the insignificance of Z2* does cast some doubt over the significance of the results obtained with Z1*.

In summary, R² comparisons based on the pooled, fixed-effects model and based on the aggregated annual cross-sectional tests derived from Z1* weakly confirm results from the main test of H1.

Tests of H1a–H1d Based on R² Comparisons

Table 5 also presents the supplementary tests of H1a–H1d based on R² comparisons for various data partitions. Not surprisingly, the R² comparisons in the partitioned analysis are weaker than those based on the pooled analysis, most likely because of the lower statistical power arising from the data partitioning. The R² values are significantly different only in the high OIL_VOLATILITY partition where the present value specification's R² is greater than that of the historical cost specification's R² ($0.71 > 0.67$, $p < .001$). The higher R² in the present value specification (relative to the R² in the historical cost specification) when oil prices are volatile supports H1c, suggesting that measurement error variance is lower in present value relative to historical costs when oil prices are volatile, perhaps because investors' expectations of future oil prices are related to spot prices.

Sensitivity Analysis

I evaluate the sensitivity of the reported results by repeating the analysis after eliminating observations that standard outlier diagnostic techniques identify as potentially influential (i.e., Belsley et al. 1980). I also estimate change-form specifications of the models.¹⁰ The overall tenor of the results is similar to those reported.

⁹ Z1* equals $\bar{z}\sqrt{N}$ and Z2* equals $\bar{z}/(\text{stddev}(z)/\sqrt{(N-1)})$. Z1* assumes independence in the annual Z-statistics, whereas Z2* corrects for potential lack of independence.

¹⁰ I performed two types of change-form analysis. One involved simply first-differencing the transformed explanatory variables (i.e., the Z variables) used in estimating the models reported. This approach provides some insight into whether the measurement error results are influenced by omitted variables that are correlated with MVE1* and MVE2*. The other involved first-differencing the "raw" explanatory variables before running the first-stage regressions and then calculating the transformed explanatory variables from the first-differenced raw variables. Because changes in successive balance sheets articulate through the income statement (plus transactions with shareholders), this second approach examines whether measurement error differs between historical cost and reserve recognition earnings. Both approaches yielded similar results.

TABLE 5

Alternative Tests of Hypothesis 1 Based on Differences in the Explanatory Power of Oil and Gas Assets Measured at Present Value Versus Historical Cost—Analysis Based on Pooled Data and Data Partitioned Based on Factors Expected to Influence the Across-Firm and Across-Time Reliability of the Present Value Measure: 1984–1996

$$\begin{aligned}
 \text{Present Value Model: } MVE &= \alpha_j + \alpha_i + \gamma_{1,FC}(OTHER\_HC \times FC) + \gamma_{2,FC}(OG\_PV \times FC) \\
 &\quad + \gamma_{3,FC}(LIAB\_HC \times FC) + \gamma_{1,SE}(OTHER\_HC \times SE) \\
 &\quad + \gamma_{2,SE}(OG\_PV \times SE) + \gamma_{3,SE}(LIAB\_HC \times SE) + \varepsilon_1 \\
 \text{Historical Cost Model: } MVE &= \alpha_j + \alpha_i + \gamma_{1,FC}(OTHER\_HC \times FC) + \gamma_{2,FC}(OG\_HC \times FC) \\
 &\quad + \gamma_{3,FC}(LIAB\_HC \times FC) + \gamma_{1,SE}(OTHER\_HC \times SE) \\
 &\quad + \gamma_{2,SE}(OG\_HC \times SE) + \gamma_{3,SE}(LIAB\_HC \times SE) + \varepsilon_2^a
 \end{aligned}$$

	<i>Present Value Model</i>	<i>Historical Cost Model</i>	<i>Significance Levels from Vuong Tests for Equal R²s</i>
<i>Test of H1</i>			
R ² from a pooled fixed-effects regression ^b	0.68	0.66	0.08
Mean R ² from annual cross-sectional regressions ^c	0.36	0.32	0.06 ^d 0.12 ^e
<i>Test of H1a</i>			
R ² from a pooled fixed-effects regression that includes only firm-years with EST_ERROR below pooled median	0.73	0.71	0.19
R ² from a pooled fixed-effects regression that includes only firm-years with EST_ERROR at or above pooled median	0.63	0.61	0.37
<i>Test of H1b</i>			
R ² from a pooled fixed-effects regression that includes only firm-years with INSIDE = 0	0.67	0.65	0.17
R ² from a pooled fixed-effects regression that includes only firm-years with INSIDE = 1	0.83	0.82	0.57
<i>Test of H1c</i>			
R ² from a pooled fixed-effects regression that includes only firm-years with OIL_VOLATILITY below pooled median	0.69	0.69	0.80
R ² from a pooled fixed-effects regression that includes only firm-years with OIL_VOLATILITY at or above pooled median	0.71	0.67	0.00
<i>Test of H1d</i>			
R ² from a pooled fixed-effects regression that includes only firm-years with RATE below pooled median	0.67	0.65	0.18

(Continued on next page)

TABLE 5 (Continued)

	<i>Present Value Model</i>	<i>Historical Cost Model</i>	<i>Significance Levels from Vuong Tests for Equal R²s</i>
R ² from a pooled fixed effects regression that includes only firm-years with RATE at or above pooled median	0.73	0.71	0.27

*, **, *** Denote significance at the 0.10, 0.05, and 0.01 levels, respectively, in a two-tailed test.

^a The parameters α_j and α_t denote firm and year fixed effects. FC and SE are dummy variables assigned a value of 1 if the firm uses the full cost or successful efforts method, respectively, and 0 otherwise. All other variables are defined in Table 1.

^b The pooled fixed-effects specification corresponds to unrestricted Model 2 in Table 4.

^c The annual cross-sectional specification corresponds to unrestricted Models 4–8 in Table 4.

^d Test of whether the time-series mean Vuong z-statistic equals 0, based on $Z1^* = \bar{z}\sqrt{N}$.

^e Test of whether the time-series mean Vuong z-statistic equals 0, based on $Z2^* = \bar{z}/\text{stddev}(z)/\sqrt{(N-1)}$.

To assess whether the inferences are sensitive to goodwill and other unrecognized assets that could potentially represent correlated omitted variables, I probe for possible effects of unrecognized assets. Harris and Ohlson (1987) argue that aggressive exploration creates goodwill. Because firms with aggressive exploration effort replace a larger percentage of reserves, I use the percentage of reserves added in the current year as a proxy for goodwill. I repeat the analysis adding this variable as another regressor, and the inferences drawn from the tests are unaffected.

In two-stage estimation, one typically uses parameters estimated in the first-stage regression to generate regressors used in the second-stage regression. Because the parameters estimated in the first-stage regression are themselves random variables rather than fixed values, one risks understating the second-stage standard errors if in the second stage the values of the generated regressors are treated as fixed values rather than as random variables estimated with error (Pagan 1984; Murphy and Topel 1985). The results I report treat the first-stage estimates as fixed values; therefore, the second-stage standard errors potentially are understated. To assess whether any such understatement influences my inferences, I use the procedure outlined in Murphy and Topel (1985) to adjust the second-stage standard errors and repeat the test of H1. Intuitively, the Murphy and Topel (1985) adjustment transforms the second-stage standard errors by a function of the first-stage standard errors. The second-stage-adjusted standard errors are indeed larger, but inferences are qualitatively unchanged, remaining significant at the 0.01 level or stronger.

Finally, estimation error in the deflator BOE could bias the measurement error tests. Both BOE and OG_PV are functions of reserve quantities, so any estimation error in reserve quantities affects both measures. Deflating oil and gas asset present values by barrel-of-oil equivalent units offsets errors in reserve quantity estimates in both the numerator and denominator, so the deflated present value measure of oil and gas assets may manifest *less* measurement error variance than the undeflated measure. OG_HC, not a function of reserve quantities, is unaffected by error in estimating reserve quantities, so when deflated it may exhibit *more* measurement error variance. That is, deflating by barrel-of-oil equivalent units may dampen measurement error variance in deflated oil and gas asset present values but amplify it in deflated oil and gas asset historical costs. I address this concern in three ways.

First, I purge barrel-of-oil equivalent units of measurement error by regressing it on oil and gas sales revenue (a size variable highly correlated with barrel-of-oil equivalent units but unlikely to have substantial measurement error), retain the resulting fitted values, and repeat the analyses after deflating by the fitted value of barrel-of-oil equivalent units. The results were virtually identical to those reported.

Second, I repeat the analyses using undeflated variables and base statistical inferences on a heteroskedasticity-consistent covariance matrix, because diagnostic tests indicate heteroskedasticity (White 1980). The undeflated analysis led to similar but significantly weaker results than the deflated analysis. I interpret the weaker significance levels as attributable to loss in estimation efficiency arising from the heteroskedasticity in the undeflated model.

Third, I simulate the effect of measurement error in barrel-of-oil equivalent units on the Type I error rate in a test of measurement error variance in *deflated* oil and gas asset present values and historical costs when the null hypothesis of equal variances is true in the *undeflated* variables. Simulation results indicate that for plausible ranges of measurement error in the deflator and the oil and gas asset measures, deflating inflates the Type I error rate by only 0.30 of 1 percent due to a simultaneous dampening of measurement error in deflated oil and gas asset present values and an amplifying of measurement error in deflated oil and gas asset historical costs.¹¹ Based on (1) the instrumental variable analysis, (2) the similarity of deflated and undeflated results, and (3) the simulation analysis, I conclude that noise in the deflator is unlikely to have influenced my inferences.

VI. INTERPRETATION AND CONCLUDING COMMENTS

I am able to replicate the findings of prior research when I estimate annual, imputed value models similar to the models these prior studies used (Magliolo 1986; Harris and Ohlson 1987; Shaw and Wier 1993). In particular, during 1984–88 (the years common to both my study and prior studies) the present value measure of oil and gas assets exhibits significantly *less* explanatory power for stock prices than the corresponding historical cost measure when I use the restricted model specifications from prior research. However, specification tests show that these annual, imputed value models are misspecified, and reject this approach in favor of an unrestricted, fixed-effects balance sheet valuation model. In the unrestricted, fixed-effects model, oil and gas assets measured at present value exhibit significantly *more* explanatory power for stock prices than the corresponding historical cost measure, both in the 1984–1996 full period and in the 1984–88 subperiod.

Consistent with the valuation analysis, I find that measurement error variance in the present value measure of oil and gas assets is less than in the corresponding historical cost measure, in both the 1984–1996 full period and the 1984–88 subperiod that corresponds to the period examined in prior research. Analysis of across-firm and across-time variation in present value measurement error variance reveals that the measurement error increases in full cost firms as revisions in reserve quantity estimates increase and as firm-specific discount rates deviate from the uniform 10 percent rate that SFAS No. 69 requires. There is weak evidence that present value measurement error variance decreases when oil price volatility is high, and no evidence that present value measurement error variance is different when firms employ an external petroleum engineer to estimate reserve volumes rather than relying on their own internal engineers.

¹¹ The intuition behind this result is simple. If oil reserves estimates are off by 1 barrel of oil, then this will induce 1 unit of measurement error in BOE but will induce $(p - x)/(1 + .10)^n$ units of measurement error in OG_PV, where p is the current price of oil, x is the current cost of extracting that barrel of oil, and n is the number of years until that barrel will be extracted. In most cases, $(p - x)/(1 + .10)^n$ substantially exceeds 1. Thus, the dampening of measurement error in deflated OG_PV is minor.

Overall, the evidence suggests that model misspecification, rather than measurement error or time-period idiosyncrasy, most likely explains the weak value relevance of present value measures of oil and gas assets reported in prior research. My findings are important because (1) they help resolve an anomaly in the literature, and (2) they demonstrate that present value offers a viable means of obtaining reliable fair value estimates for oil and gas assets—nonfinancial assets that have little firm-specific value and are traded in active markets. Of course, I cannot rule out that sampling difference contributes to the differences between my results and those reported in prior research, nor can I eliminate the possibility that time-period idiosyncratic factors unique to the period 1979–1983 (i.e., the period examined in prior research that I do not examine) accounts for the weak association between stock prices and the oil and gas asset present value measure as reported in Magliolo (1986) and Harris and Ohlson (1987).

One should interpret the results in this paper in light of the study's limitations. First, the measurement error analysis assumes that investors' implicit assessed value of assets and liabilities provides a meaningful benchmark for assessing measurement error. If stock prices deviate from "intrinsic value," then investor-assessed values provide noisy benchmarks that, in turn, could affect my inferences. Second, because my selection criteria limit the sample to large firms (i.e., firms on Compustat and CRSP) that can arguably devote more expertise and resources to preparing present value estimates, it is not clear whether the present value measure of oil and gas assets would have greater explanatory power and less measurement error in a sample of smaller firms. Finally, I interpret the measurement error tests under a maintained hypothesis about certain aspects of the measurement error structure as detailed in the Appendix. Although the assumptions that form the maintained hypothesis follow logically from an analysis of differences in recognition and measurement under reserve recognition vs. historical cost accounting for oil and gas operations, I cannot test the descriptive validity of these assumptions.

APPENDIX

MEASUREMENT ERROR MODEL DETAILS

Financial statements and related disclosures measure the market values of assets and liabilities with error as Equation (2) in the body of the paper shows:

$$\begin{aligned}\text{OTHER\_HC} &= \text{OTHER\_MV} + u_1 \\ \text{OG\_i} &= S_i \times (\text{OG\_MV} + u_{2i}) \\ \text{LIAB\_HC} &= \text{LIAB\_MV} + u_3\end{aligned}\tag{2}$$

Each error term, u_x , is a random variable with variance σ_x^2 . Covariance between x and y is denoted $\sigma_{x,y}$. S_i is a parameter reflecting the difference in scale between OG_HC and OG_PV. Measurement error as defined by Equation (2) arises when accounting values and market values become misaligned.

Rao (1973), Garber and Klepper (1980), Barth (1989, 1991) and Choi et al. (1997) show that measurement error in least squares regressors causes biased coefficient estimates. Their findings suggest that a least squares regression of the theoretical valuation model (specified as Equation [1] in the body of the paper) estimated using measured-with-error accounting values as proxies for the unobservable market values will yield coefficient bias as specified in Equations (A1) and (A2) when OG_MV is proxied by OG_PV and OG_HC, respectively.

$$\begin{aligned}
\text{MVE} = & \gamma_0 + \left[1 - \underbrace{\sum_{\substack{i=1 \\ i \neq 1}}^3 \frac{\alpha_{i1}\phi_i}{\sigma_{e_i}^2} + \frac{\phi_1}{\sigma_{e_1}^2}}_{\text{Bias}} \right] \text{OTHER\_HC} \\
& + \left[1 - \underbrace{\sum_{\substack{i=1 \\ i \neq 2}}^3 \frac{-\alpha_{i2}\phi_i}{\sigma_{e_i}^2} + \frac{\phi_2}{\sigma_{e_2}^2}}_{\text{Bias}} \right] \text{OG\_PV} \\
& + \left[1 - \underbrace{\sum_{\substack{i=1 \\ i \neq 3}}^3 \frac{-\alpha_{i3}\phi_i}{\sigma_{e_i}^2} + \frac{\phi_3}{\sigma_{e_3}^2}}_{\text{Bias}} \right] \text{LIAB\_HC} + \varepsilon_1
\end{aligned} \tag{A1}$$

$$\begin{aligned}
\text{MVE} = & \gamma_0 + \left[1 - \underbrace{\sum_{\substack{i=1 \\ i \neq 1}}^3 \frac{-\beta_{i1}\delta_i}{\sigma_{e_i}^{*2}} + \frac{\delta_1}{\sigma_{e_1}^{*2}}}_{\text{Bias}} \right] \text{OTHER\_HC} \\
& + \left[1 - \underbrace{\sum_{\substack{i=1 \\ i \neq 2}}^3 \frac{-\beta_{i2}\delta_i}{\sigma_{e_i}^{*2}} + \frac{\delta_2}{\sigma_{e_2}^{*2}}}_{\text{Bias}} \right] \text{OG\_HC} \\
& + \left[1 - \underbrace{\sum_{\substack{i=1 \\ i \neq 3}}^3 \frac{-\beta_{i3}\delta_i}{\sigma_{e_i}^{*2}} + \frac{\delta_3}{\sigma_{e_3}^{*2}}}_{\text{Bias}} \right] \text{LIAB\_HC} + \varepsilon_2
\end{aligned} \tag{A2}$$

ϕ_i in Equation (A1) and δ_i in Equation (A2) measure the variances/covariances among the measurement error terms (u_1 , u_{2i} , and u_3). α_{ij} in Equation (A1) measures the partial correlations among the accounting measures when OG_PV proxies for OG_MV, and is operationally defined as the slope coefficients obtained from the regression specified in Equation (A3):

$$\begin{aligned}
\text{OTHER\_HC} &= \alpha_{10} + \alpha_{12}\text{OG\_PV} + \alpha_{13}\text{LIAB\_HC} + e_1 \\
\text{OG\_PV} &= \alpha_{20} + \alpha_{21}\text{OTHER\_HC} + \alpha_{23}\text{LIAB\_HC} + e_2 \\
\text{LIAB\_HC} &= \alpha_{30} + \alpha_{31}\text{OTHER\_HC} + \alpha_{32}\text{OG\_PV} + e_3
\end{aligned} \tag{A3}$$

β_{ij} measures the partial correlations among the accounting measures when OG_HC proxies for OG_MV and is operationally defined as the slope coefficients obtained from the regression in Equation (A4):

$$\begin{aligned}
\text{OTHER\_HC} &= \beta_{10} + \beta_{12}\text{OG\_HC} + \beta_{13}\text{LIAB\_HC} + e_1^* \\
\text{OG\_HC} &= \beta_{20} + \beta_{21}\text{OTHER\_HC} + \beta_{23}\text{LIAB\_HC} + e_2^* \\
\text{LIAB\_HC} &= \beta_{30} + \beta_{31}\text{OTHER\_HC} + \beta_{32}\text{OG\_HC} + e_3^*
\end{aligned} \tag{A4}$$

Choi et al. (1997) show that Equations (A1) and (A2) can be re-expressed as shown in Equations (A5) and (A6), respectively:

$$\text{MVE1}^* = \phi_0 + \phi_1 Z1_{\text{OTHER\_HC}} + \phi_2 Z1_{\text{OG\_PV}} + \phi_3 Z1_{\text{LIAB\_HC}} + \varepsilon_1 \tag{A5}$$

$$\text{MVE2}^* = \delta_0 + \delta_1 Z2_{\text{OTHER\_HC}} + \delta_2 Z2_{\text{OG\_HC}} + \delta_3 Z2_{\text{LIAB\_HC}} + \varepsilon_2 \tag{A6}$$

where:

$$\begin{aligned}
Z1_{\text{OTHER\_HC}} &= -\left(\frac{\text{OTHER\_HC} - \alpha_{12}\text{OG\_PV} - \alpha_{13}\text{LIAB\_HC}}{\sigma_{e_1}^2}\right) \\
Z1_{\text{OG\_PV}} &= -\left(\frac{\text{OG\_PV} - \alpha_{21}\text{OTHER\_HC} - \alpha_{23}\text{LIAB\_HC}}{\sigma_{e_2}^2}\right) \\
Z1_{\text{LIAB\_HC}} &= -\left(\frac{\text{LIAB\_HC} - \alpha_{31}\text{OTHER\_HC} - \alpha_{32}\text{OG\_PV}}{\sigma_{e_3}^2}\right) \\
Z2_{\text{OTHER\_HC}} &= -\left(\frac{\text{OTHER\_HC} - \beta_{12}\text{OG\_HC} - \beta_{13}\text{LIAB\_HC}}{\sigma_{e_1^*}^2}\right) \\
Z2_{\text{OG\_HC}} &= -\left(\frac{\text{OG\_HC} - \beta_{21}\text{OTHER\_HC} - \beta_{23}\text{LIAB\_HC}}{\sigma_{e_2^*}^2}\right) \\
Z2_{\text{LIAB\_HC}} &= -\left(\frac{\text{LIAB\_HC} - \beta_{31}\text{OTHER\_HC} - \beta_{32}\text{OG\_HC}}{\sigma_{e_3^*}^2}\right)
\end{aligned} \tag{A7}$$

MVE1* equals MVE - (OTHER_HC + OG_PV + LIAB_HC) and MVE2* equals MVE - (OTHER_HC + OG_HC + LIAB_HC).

Subtracting from δ_i from ϕ_i cancels common covariance terms, yielding:

$$\begin{aligned}
\phi_1 - \delta_1 &= \sigma_{u_1, u_{2PV}} - \sigma_{u_1, u_{2HC}} + \sigma_{\text{OTHER\_MV}, u_{2PV}} - \sigma_{\text{OTHER\_MV}, u_{2HC}} \\
\phi_2 - \delta_2 &= \sigma_{u_{2PV}}^2 - \sigma_{u_{2HC}}^2 + \sigma_{u_1, u_{2PV}} - \sigma_{u_1, u_{2HC}} + \sigma_{u_{2PV}, u_3} - \sigma_{u_{2HC}, u_3} + \sigma_{\text{OG\_MV}, u_{2PV}} \\
&\quad - \sigma_{\text{OG\_MV}, u_{2HC}} \\
\phi_3 - \delta_3 &= \sigma_{u_{2PV}, u_3} - \sigma_{u_{2HC}, u_3} + \sigma_{\text{LIAB\_MV}, u_{2PV}} - \sigma_{\text{LIAB\_MV}, u_{2HC}}
\end{aligned} \tag{A8}$$

Rearranging Equation (A8) yields Equation (5) in the body of the paper:

$$\begin{aligned}
 \delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3 = & \sigma_{u2PV}^2 - \sigma_{u2HC}^2 + \underbrace{\sigma_{OTHER\_MV,u2HC} - \sigma_{OTHER\_MV,u2PV}}_{(-)} \\
 & + \underbrace{\sigma_{OG\_MV,u2PV} - \sigma_{OG\_MV,u2HC}}_{(+)} \\
 & + \underbrace{\sigma_{LIAB\_MV,u2HC} - \sigma_{LIAB\_MV,u2PV}}_{(+)}
 \end{aligned} \tag{5}$$

The parentheses below the covariance difference terms in Equation (5) identify the expected sign of the expression based on the posited sign of and relationship among the covariance difference terms as discussed next.

Because measurement error is defined as misalignment between accounting and market values, any factor causing market values to increase (decrease) relative to accounting values also creates a negative (positive) realized value for the measurement error term. Thus, by definition, measurement error terms and market values must move in opposite directions, implying negative covariance, or $\sigma_{OTHER\_MV,u1} < 0$, $\sigma_{OG\_MV,u2i} < 0$, and $\sigma_{LIAB\_MV,u3} < 0$. Furthermore, the measurement error terms will not be independent, because interest rate shocks should trigger simultaneous misalignment between accounting and market values of all assets and liabilities. Because, first, measurement error terms are not independent and, second, measurement error and market values have negative covariance, it follows that measurement error in oil and gas assets will have negative covariance with *OTHER_MV* and positive covariance with *LIAB_MV*, or $\sigma_{OTHER\_MV,u2i} < 0$ and $\sigma_{LIAB\_MV,u2i} > 0$.

Moreover, because the same set of common factors (changes in interest rates and expected cash flows) creates measurement error in *OTHER_HC*, *LIAB_HC*, and *OG_HC* whereas only one of these factors (changes in interest rates) creates measurement error in *OG_PV*, measurement error in *OG_HC* as compared to *OG_PV* should be more strongly associated with measurement error in *OTHER_HC* and *LIAB_HC*, and their market values. Consequently, measurement error covariance in *OG_PV* ($\sigma_{x,u2PV}$) will be of the same *sign* yet smaller in absolute value (i.e., weaker) than measurement error covariance in *OG_HC* ($\sigma_{x,u2HC}$). That is, $|\sigma_{x,u2HC}| > |\sigma_{x,u2PV}|$. Based on the chain of logic that $\sigma_{OG\_MV,u2i} < 0$, $\sigma_{OTHER\_MV,u2i} < 0$, $\sigma_{LIAB\_MV,u2i} > 0$ and $|\sigma_{x,u2HC}| > |\sigma_{x,u2PV}|$, I conclude that the covariance difference terms in Equation (5) will have the following signed values: $\sigma_{OTHER\_MV,u2HC} - \sigma_{OTHER\_MV,u2HC} < 0$, $\sigma_{OG\_MV,u2PV} - \sigma_{OG\_MV,u2HC} > 0$, and $\sigma_{LIAB\_MV,u2HC} - \sigma_{LIAB\_MV,u2PV} > 0$. This line of logic shows that two of the three covariance difference terms should be positive. Additionally, because the magnitude of *OG_MV* is larger than *OTHER_MV* (because oil and gas firms invest the largest fraction of their total assets in oil and gas assets), and because covariance is a magnitude-sensitive metric, I expect the negative covariance difference term, $\sigma_{OTHER\_MV,u2HC} - \sigma_{OTHER\_MV,u2PV}$, to be smaller in absolute value than the positive covariance difference term, $\sigma_{OG\_MV,u2PV} - \sigma_{OG\_MV,u2HC}$. As a result, I expect the covariance terms in Equation (5) to net to a positive signed value.

A statistical test of H1, that $\sigma_{u2PV}^2 = \sigma_{u2HC}^2$, based on Equation (5) is: $H_0: \delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3 = 0$. Because the covariance terms in Equation (5) should net to a positive value, the coefficient function $\delta_1 - \delta_2 + \delta_3 - \phi_1 + \phi_2 - \phi_3$ is biased toward a positive value under the null hypothesis that measurement error variance in *OG_PV* and *OG_HC*

are the same (i.e., $H_0: \sigma_{u_{2PV}}^2 = \sigma_{u_{2HC}}^2$). Finally, Appendix Table A1 summarizes the covariance structure that I posit based on the measurement differences between OG_HC and OG_PV.

Control for Scale Effects

The role of scale is evident when I re-express the specification of OG_PV and OG_HC in Equation (2) as:

$$\begin{aligned} \text{OG\_PV} &= S_{PV} \times (\text{OG\_MV} + u_{2PV}) \text{ and} \\ \text{OG\_HC} &= S_{HC} \times (\text{OG\_MV} + u_{2HC}) \end{aligned} \quad (\text{A9})$$

where S_{PV} and S_{HC} are scale factors. In this revised specification, the variance of measurement error in OG_PV is $S_{PV}^2 \times \sigma_{u_{2PV}}^2$ and the variance of OG_HC is $S_{HC}^2 \times \sigma_{u_{2HC}}^2$. Thus, one cannot reliably test $H_0: \sigma_{u_{2PV}}^2 = \sigma_{u_{2HC}}^2$ unless $S_{PV} = S_{HC}$. Multiplying OG_HC by S_{PV}/S_{HC} transforms u_{2HC} to the same scale as u_{2PV} , and thus facilitates a scale-free test of $H_0: \sigma_{u_{2PV}}^2 = \sigma_{u_{2HC}}^2$. Consequently, I ran all measurement error regressions after substituting $(S_{PV}/S_{HC}) \times \text{OG\_HC}$ in place of OG_HC. I estimate S_{PV}/S_{HC} by dividing the mean of OG_PV by the mean OG_HC.

TABLE A1
Posited Signs of and Relations among Covariances

<i>Covariance</i>	<i>Reasoning</i>
$\sigma_{\text{OTHER\_MV}, u_1} < 0$ $\sigma_{\text{OG\_MV}, u_{2i}} < 0$ $\text{LIAB\_MV}, u_3 < 0$	Measurement error is defined as misalignment between accounting and market values. Any factor causing market values to increase (decrease) relative to accounting values creates a negative (positive) realized measurement error. Thus, by definition, measurement error terms and market values must move in opposite directions, implying negative covariance.
$\sigma_{u_1 u_{2i}} > 0$ $\sigma_{u_1 u_3} < 0$ $\sigma_{u_{2i} u_3} < 0$	Asset and liability values are sensitive to similar factors, implying a positive covariance between asset market values and a negative covariance between asset market values and liability market values (the covariance is negative since liabilities are defined as negative values). Since these changes in asset and liability values cause changes of the opposite sign in their related accounting measurement error terms, the error terms u_1 and u_{2i} will exhibit positive covariance while u_3 will exhibit negative covariance with u_1 and u_{2i} .
$\sigma_{\text{OTHER\_MV}, u_{2i}} < 0$ $\sigma_{\text{OG\_MV}, u_1} < 0$	Implied by the positive covariance in asset measurement error terms ($\sigma_{u_1 u_{2i}} > 0$) and the negative covariance with asset values ($\sigma_{\text{OTHER\_MV}, u_1} < 0$ and $\sigma_{\text{OG\_MV}, u_{2i}} < 0$).
$\sigma_{\text{LIAB\_MV}, u_1} > 0$ $\sigma_{\text{LIAB\_MV}, u_{2i}} > 0$	Implied by the negative covariance between asset and liability measurement error ($\sigma_{u_1 u_3} < 0$ and $\sigma_{u_{2i} u_3} < 0$) and the negative covariance between liability values and measurement error ($\sigma_{\text{LIAB\_MV}, u_3} < 0$).
$\sigma_{\text{OTHER\_MV}, u_3} > 0$ $\sigma_{\text{OG\_MV}, u_3} > 0$	Implied by the negative covariance between asset measurement error ($\sigma_{\text{OTHER\_MV}, u_1} < 0$ and $\sigma_{\text{OG\_MV}, u_{2i}} < 0$) and the negative covariance in asset and liability measurement error ($\sigma_{u_1 u_3} < 0$ and $\sigma_{u_{2i} u_3} < 0$).

(Continued on next page)

TABLE A1 (Continued)

Covariance	Reasoning
$\sigma_{u_1u_2HC} > \sigma_{u_1u_2PV}$ σ_{OTHER_MV,u_2HC} $< \sigma_{OTHER\_MV,u_2PV}$ σ_{OG_MV,u_2HC} $< \sigma_{OG\_MV,u_2PV}$	OG_PV is updated at each balance sheet date to reflect changing expectations of future cash flows. OTHER_HC and OG_HC are not similarly updated. The same set of common factors (changes in interest rates and changes in expected future cash flows) creates measurement error in both OTHER_HC and OG_HC, while only one of these common factors (changes in interest rates) creates measurement error in OG_PV. Thus, I expect the positive covariance between measurement error in OTHER_HC and OG_HC to be stronger than the positive covariance of measurement error between OTHER_HC and OG_PV. Similar reasoning leads to the conclusion that asset market values covary more negatively with the measurement error in OG_HC than with the measurement error in OG_PV.
$\sigma_{u_2HCu_3} < \sigma_{u_2PVu_3}$ σ_{LIAB_MV,u_2HC} $> \sigma_{LIAB\_MV,u_2PV}$	Same reasoning as above except the negative covariance between u_{2i} and u_3 and the positive covariance between LIAB_MV and u_{2i} will be weaker in OG_PV than in OG_HC.

REFERENCES

- Aboody, D., M. Barth, and R. Kasznik. 1999. Revaluations of fixed assets and firm performance: Evidence from the U.K. *Journal of Accounting and Economics* 26 (January): 149–178.
- Adams, R., D. Bach, K. Bickett, J. Carroll, R. Gallun, and P. Kuiper. 1994. The full cost ceiling test. *Journal of Accountancy* 177 (April): 67–70.
- Alciatore, M. 1993. New evidence on SFAS No. 69 and the components of the change in reserve value. *The Accounting Review* 68 (October): 639–656.
- Amir, E., T. Harris, and E. Venuti. 1993. A comparison of the value-relevance of U.S. versus non-U.S. GAAP accounting measures using Form 20-F reconciliations. *Journal of Accounting Research* 31 (Supplement): 230–264.
- Barth, M. 1989. Assessing financial accounting measurement alternatives for assets and liabilities. Ph.D. dissertation, Stanford University Graduate School of Business.
- . 1991. Relative measurement errors among alternative pension asset and liability measures. *The Accounting Review* 66 (July): 433–463.
- . 1994. Fair value accounting: Evidence from investment securities and the market valuation of banks. *The Accounting Review* 69 (January): 1–25.
- , and W. Landsman. 1995. Fundamental issues related to using fair value accounting for financial reporting. *Accounting Horizons* 9 (December): 97–107.
- , W. Beaver, and W. Landsman. 1996. Value-relevance of banks' fair value disclosures under SFAS No. 107. *The Accounting Review* 71 (October): 513–538.
- , and G. Clinch. 1998. Revalued financial, tangible, and intangible assets: Associations with share prices and non-market-based value estimates. *Journal of Accounting Research* 36 (Supplement): 199–233.
- Belsley, D., E. Kuh, and R. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York, NY: Wiley.
- Choi, B., D. Collins, and W. Johnson. 1997. Valuation implications of reliability differences: The case of nonpension postretirement obligations. *The Accounting Review* 72 (July): 351–383.
- Christie, A. 1990. Aggregation of test statistics: An evaluation of the evidence on contracting and size hypotheses. *Journal of Accounting and Economics* 12 (January): 15–36.
- Clinch, G., and J. Magliolo. 1992. Market perceptions of reserve disclosures under SFAS No. 69. *The Accounting Review* 67 (4): 843–861.

- Deakin, E. 1979. An analysis of differences between non-major oil firms using successful efforts and full cost methods. *The Accounting Review* 54 (October): 722–734.
- Doran, B., D. Collins, and D. Dhaliwal. 1988. The information of historical cost earnings relative to supplemental reserve-based accounting data in the extractive petroleum industry. *The Accounting Review* 63 (3): 389–413.
- Easton, P., P. Edey, and T. Harris. 1993. An investigation of revaluations of tangible long-lived assets. *Journal of Accounting Research* 31 (Supplement): 1–38.
- Eccher, E., K. Ramesh, and S. Thiagarajan. 1996. Fair value disclosures of bank holding companies. *Journal of Accounting and Economics* 22 (August/December): 79–117.
- Financial Accounting Standards Board (FASB). 1982. *Disclosures about Oil and Gas Producing Activities*. Statement of Financial Accounting Standards No. 69. Stamford, CT: FASB.
- . 2000. *Using Cash Flow Information and Present Value in Accounting Measurements*. Statement of Financial Accounting Concepts No. 7. Norwalk, CT: FASB.
- Garber, S., and S. Klepper. 1980. “Administer pricing” or competition coupled with measurement? *International Economic Review* 21 (June): 413–435.
- Hall, S., and W. Stammerjohan. 1997. Damage awards and earnings management in the oil industry. *The Accounting Review* 72 (1): 47–65.
- Harris, T., and J. Ohlson. 1987. Accounting disclosures and the market’s valuation of oil and gas properties. *The Accounting Review* 62 (October): 651–670.
- Healy, P., S. Kang, and K. Palepu. 1987. The effect of accounting procedure changes on CEOs’ cash salary and bonus compensation. *Journal of Accounting and Economics* 9 (April): 7–34.
- Kahn, N., J. Krausz, and A. Schiff. 1983. Another view of the reliability of oil and gas reserve estimates. *Journal of Extractive Industries Accounting* 2 (Summer): 103–115.
- King, B. 1982. Oil and gas disclosures: Some empirical results. *Journal of Extractive Industries Accounting* 1 (Spring): 107–127.
- Landsman, W. 1986. An empirical investigation of pension fund property rights. *The Accounting Review* 61 (October): 662–691.
- Lys, T. 1986. Discussion of capital analysis of reserve recognition accounting. *Journal of Accounting Research* 24 (Supplement): 109–111.
- Magliolo, J. 1986. Capital market analysis of reserve recognition accounting. *Journal of Accounting Research* 24 (Supplement): 69–108.
- Malmquist, D. 1990. Efficient contracting and the choice of accounting method in the oil and gas industry. *Journal of Accounting and Economics* 12 (January): 173–205.
- Murphy, K., and R. Topel. 1985. Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics* 3 (October): 370–379.
- Nelson, K. 1996. Fair value accounting for commercial banks: An empirical analysis of SFAS No. 107. *The Accounting Review* 71 (April): 161–182.
- Pagan, A. 1984. Econometric issues in the analysis of regressions with generated regressors. *International Economic Review* 25 (February): 221–247.
- Porter, S. 1980. *Highlights of a Study of the Subjectivity of Reserve Estimates and Its Relation to Financial Reporting*. Tulsa, OK: Arthur Young and Company.
- Rao, P. 1973. Some notes on the errors-in-variables model. *The American Statistician* 27 (December): 217–218.
- Shaw, W., and H. Wier. 1993. Organizational form choice and the valuation of oil and gas producers. *The Accounting Review* 68 (3): 657–667.
- U.S. Securities and Exchange Commission (SEC). 1978. *Accounting Series Release No. 253*. Washington, D.C.: SEC.
- . 1981. *Accounting Series Release No. 289*. Washington, D.C.: SEC.
- Venkatachalam, M. 1996. Value-relevance of banks’ derivatives disclosures. *Journal of Accounting and Economics* 22 (August/December): 327–355.
- Vuong, Q. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57 (March): 307–333.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (May): 817–838.

An Exploratory Study of the Valuation Properties of Cross-Listed Firms' IAS and U.S. GAAP Earnings and Book Values

Hollis Ashbaugh

University of Wisconsin–Madison

Per Olsson

Duke University

ABSTRACT: Despite the increasing integration of global capital markets, there is little evidence on the valuation properties of cross-listed, non-U.S. firms' accounting variables. We use the relative performance of the earnings capitalization, the book value, and the residual income valuation models to explore the valuation properties of International Accounting Standards and U.S. Generally Accepted Accounting Principles earnings and book values reported by non-U.S., cross-listed firms trading in a common equity market. Using non-U.S./non-U.K. firms whose shares trade on the International Stock Exchange Automated Quotation system in London, we find that the earnings capitalization model is the dominant accounting-based valuation model when cross-listed firms report under International Accounting Standards. In contrast, we find that when cross-listed firms report under U.S. Generally Accepted Accounting Principles, the residual income model is the dominant accounting-based valuation model. Our exploratory study provides insights into the valuation implications of allowing a dual reporting system for foreign registrants trading in a common equity market.

We thank Daniel B. Thornton, the associate editor, an anonymous referee, workshop participants at The University of Iowa and the University of Wisconsin–Madison, participants at the 2000 AAA International Midyear Meeting, and Rick Antle, Dan Collins, Doug De Jong, Jennifer Francis, Ole-Kristian Hope, Dennis Oswald, Joel Pike, Mort Pincus, Wayne Thomas, Mohan Venkatachalam, and Terry Warfield for their comments. We also appreciate the contribution of I/B/E/S International Inc. for providing analyst following, available through the Institutional Brokers Estimate System.

*Submitted December 1999
Accepted May 2001*

I. INTRODUCTION

Despite the increasing integration of global capital markets, there is little evidence on the valuation properties of cross-listed, non-U.S. firms' accounting variables. International Accounting Standards (IAS) and U.S. Generally Accepted Accounting Principles (U.S. GAAP) are alternative sets of internationally acceptable financial reporting standards. The objective of both IAS and U.S. GAAP is to generate information that current and potential investors can use in making rational investment decisions, i.e., financial information that is useful in estimating the intrinsic value of a firm's shares (International Accounting Standards Committee [IASC] 1997; Financial Accounting Standards Board [FASB] 1978). Many non-U.S. equity markets already allow foreign registrants to file financial information prepared in accordance with either IAS or U.S. GAAP (IASC 2000). The U.S. Securities and Exchange Commission (SEC) is considering allowing foreign registrants to prepare SEC filings in accordance with IAS without reconciliation to U.S. GAAP (SEC 2000). The purpose of this exploratory study is to investigate the valuation properties of IAS and U.S. GAAP earnings and book values reported by non-U.S., cross-listed firms trading in a common equity market.

We use the relative performance of the earnings capitalization, book value, and residual income valuation models to draw inferences on the valuation properties of cross-listed firms' IAS and U.S. GAAP earnings and book values. We employ these three accounting-based valuation models because the models are used by investors and in academic research (Goldman Sachs 1999; Barker 1998; Collins et al. 1999). We assess each model's performance by examining its explanatory power and testing whether the estimated parameters lie within the bounds predicted by the model. We use a sample of 55 non-U.S./non-U.K. firms that report IAS or U.S. GAAP accounting variables and whose shares trade in the Stock Exchange Automated Quotations (SEAQ) International Equity Market of London, since SEAQ International is the largest international equity market (Zeff 1998). Restricting the sample to one equity market controls for market institutional features that can otherwise affect price formation and the inferences drawn from the analyses (Frost and Pownall 1994). Restricting the sample to one equity market also allows us to provide insights into the valuation implications of allowing a dual reporting system for foreign registrants in a common equity market. The disadvantage of limiting our sample to firms trading on SEAQ International is that this yields a relatively small sample, potentially limiting the strength of our conclusions. Therefore, although our study does provide insight into the financial-reporting issues associated with the growth in cross-listed firms, we view it as an exploratory study.

We first explore the performance of the earnings capitalization, book value, and residual income models using the 26 firms that report IAS earnings and book values. The FASB identifies the differences between IAS and U.S. GAAP accounting standards and suggests that IAS allows managers more discretion in their measurement and recognition methods than does U.S. GAAP (FASB 1999). The results in Ashbaugh (2001) suggest that cross-listed firms choose to report IAS rather than U.S. GAAP earnings and book values because of the additional discretion in accounting measurement and recognition methods available under IAS. Whether managers' discretion under IAS affects the valuation properties of cross-listed firms' IAS accounting variables is an unanswered question. This study is the first to document the valuation properties of cross-listed firms' IAS earnings and book values using alternative valuation models, thus providing evidence on the valuation usefulness of cross-listed firms' IAS accounting variables. This is a concern of the SEC in determining whether IAS is acceptable for U.S. filing requirements (SEC 2000).

Next we explore the performance of the accounting-based valuation models using the 36 non-U.S., cross-listed firms that report under U.S. GAAP. Although recent research suggests that the residual income model is the dominant valuation model for U.S. firms that report in accordance with U.S. GAAP (Penman and Sougiannis 1998; Francis et al. 2000), prior research has not investigated the performance of the residual income model when valuing non-U.S. firms reporting under U.S. GAAP. Ball et al. (2000) argue that the extent to which formal standards determine accounting practice varies internationally, and that firms' incentives to comply with accounting standards depend on the penalties assessed for noncompliance. Frost and Kinney (1996) document substantial non-compliance with SEC reconciliation requirements in their study of U.S. foreign registrants, suggesting that U.S. foreign registrants view the costs of complying with SEC reporting requirements to exceed the expected costs of noncompliance. Thus, non-U.S. firms may implement U.S. GAAP differently than U.S. firms, which in turn could affect the valuation properties of the reported accounting variables, and, as such, the performance of the valuation models. We extend the research studying U.S. GAAP by documenting the dominant valuation model for non-U.S., cross-listed firms reporting under U.S. GAAP. This is important given the substantial increase in U.S. foreign listings over the past 10 years (New York Stock Exchange [NYSE] 1999).

Results of the IAS analyses are as follows. We find the earnings capitalization model to be the dominant valuation model when valuing IAS firms' shares. The book value and residual income models do not perform as well as the earnings capitalization model in that the explanatory power of the models is weaker and the magnitude of the coefficient on residual income in the residual income model indicates model misspecification. We identify discretionary accounting measurement and recognition methods under IAS that violate the residual income model assumptions (e.g., more than half of the IAS sample firms voluntarily upward revalue their fixed assets, a choice available under IAS No. 16, *Property, Plant, and Equipment* [IASC 1997]), thus providing an explanation for the residual income model misspecification. At the same time, we identify measurement and recognition methods under IAS (e.g., the capitalization of development costs under IAS No. 9, *Research and Development Costs* [IASC 1997]) that likely contribute to the dominance of the earnings capitalization model for IAS firms.

Turning to the U.S. GAAP analyses, we find the residual income model to be the dominant valuation model for cross-listed firms reporting under U.S. GAAP. The residual income model appears well specified, whereas the estimated parameters of the earnings capitalization and book value models indicate model misspecification. Our results are similar to those of prior research that document the performance of the three accounting-based valuation models for U.S. firms reporting in accordance with U.S. GAAP (e.g., Francis and Schipper 1999). The results suggest that non-U.S., cross-listed firms' U.S. GAAP summary accounting variables have similar valuation implications to those of U.S. firms. Our results thus indicate that U.S. GAAP implementation differences between U.S. and non-U.S. firms have little, if any, importance for investors. However, it is not clear whether these results generalize to firms not listed on multiple exchanges.

Recently, the SEC has expressed concern about the information sufficiency of U.S. foreign registrants' U.S. GAAP reconciliations (SEC 2000). Non-U.S. firms have a choice in how they report U.S. GAAP earnings and book values. They can (1) prepare a U.S. GAAP reconciliation, or (2) prepare a complete set of U.S. GAAP financial statements. We conduct an additional exploratory analysis to investigate whether non-U.S. firms' U.S. GAAP reporting method affects the performance of the three valuation models. We partition

our U.S. GAAP sample firms into U.S. GAAP reconcilers or U.S. GAAP reporters, and find that the rankings of the residual income, book value, and earnings capitalization models hold for both reporting methods. However, the explanatory power of U.S. GAAP reporters' accounting variables exceeds that of reconcilers, regardless of valuation model. The invariance of the ranking suggests that the residual income model works because of the articulation between earnings and book values under U.S. GAAP. At the same time, the results suggest that the additional disclosures inherent in U.S. GAAP reports provide additional valuation-relevant information.

The rest of the paper is organized as follows. Section II presents the three valuation models and how we expect the models to perform under IAS vs. U.S. GAAP. Section III outlines the research design. Results are presented in Section IV. The study concludes with a discussion of the results in Section V.

II. BACKGROUND AND PRIOR RESEARCH

We use the relative performance of three accounting-based valuation models, the earnings capitalization (EC), book value (BV), and residual income (RI) models, to draw inferences on the valuation properties of IAS and U.S. GAAP earnings and book values. There is a rich literature on the appropriateness of these models when valuing U.S. firms' U.S. GAAP accounting variables (e.g., Collins et al. 1999), and investors and other market participants use the EC, BV, and RI valuation models in making investment decisions (Barker 1998; Goldman Sachs 1999). Furthermore, from a research perspective, it is important to benchmark the performance of accounting-based valuation models before using such models to assess the implications of changes in accounting standards (Holthausen and Watts 2000).

All valuation models build on the principle that a firm's market price is equal to investors' expected discounted future cash flows. The EC, BV, and RI valuation models use current period accounting data to proxy for expected discounted future cash flows. We assess each model's performance by examining the ability of the respective accounting variable(s) to explain price and by testing certain parameter assumptions. Since our interest lies in the relation between current period accounting variables and price, we do not add information (e.g., analysts' forecasts) or manipulate reported accounting variables to enhance the performance of the models,¹ which we define as:

$$(EC) \quad P_i = \alpha_E + \beta_E E_i + \varepsilon_i$$

$$(BV) \quad P_i = \alpha_{BV} + \beta_{BV} BV_i + v_i$$

$$(RI) \quad P_i = \alpha_{RI} + \beta_{RI} BV_i + \gamma_{RI} RI_i + \xi_i$$

where P_i is market price per share of firm i , and E_i , BV_i and RI_i are firm i 's earnings, book value of stockholders' equity, and residual income per share, respectively.

¹ Some value-relevance studies do not explicitly state the valuation model, but a valuation model is implicitly assumed. Easton and Harris (1991) show that a regression of stock returns on U.S. GAAP earnings levels, a regression commonly used in value relevance studies, is conceptually equivalent to a regression of price on book value (i.e., the BV model). A regression of stock returns on earnings changes equals a regression of price on earnings (i.e., the EC model). We could assess the capital market effects of the differences between the two accounting regimes by using a regression of returns on levels of and changes in earnings. We opted for the price specifications because they allow us to interpret coefficients, i.e., to assess model misspecification more directly. This is an important consideration when determining the value relevance of an accounting measurement by regression analysis.

The regression model's R^2 measures the strength of the association between the accounting variable(s) and price. If the EC model is well specified, then α_E should not be statistically different from 0. A non-zero intercept in the BV model indicates that there exists assets or liabilities valued by the market that are not reported on firms' balance sheets. An estimated coefficient different from 1 in the BV model indicates that the accounting standards' measure of firms' net assets is not equal to the market's valuation of those assets.

The RI model, which assumes clean surplus accounting, is theoretically defined as $V_{i,t} = BV_{i,t} + \sum_{\tau=t+1}^{\infty} (RI_{i,\tau} / (1 + r_i)^\tau)$, where $V_{i,t}$ is the intrinsic value of firm i at valuation date t , $BV_{i,t}$ is firm i 's book value of equity at valuation date t , r_i is the expected discount rate for firm i , and $RI_{i,\tau}$ is the expected residual income of firm i for period τ (where residual income is defined as period τ net income minus the discount rate times book value at $\tau - 1$). Ohlson's (1995) analytical work and empirical research on U.S. GAAP data (e.g., Nissim and Penman 1999; Francis et al. 2001) show that residual income tends to follow a mean-reverting process. As a consequence, we can express the expected discounted residual income (the summation term) as a constant times current residual income, and write the model entirely in terms of observable accounting variables: $V_{i,t} = c_1 + BV_{i,t} + c_2 RI_{i,t}$, where c_1 and c_2 are constants, and $RI_{i,t}$ equals $E_{i,t} - r_i BV_{i,t-1}$. Note that our single-period RI model is distinct from a multiperiod RI model, as the latter requires forecasts of accounting data. If the data conform to the assumption of the single-period RI model, then γ_{RI} has an upper bound of $1/r$.² Myers (1999) shows that if accounting measurements are conservative, then the intercept in the RI model is expected to be positive. Accounting conservatism, however, does not violate the model assumptions.

Based on prior research using U.S. firms, we expect the RI model to be the dominant valuation model when non-U.S., cross-listed firms report U.S. GAAP accounting variables. However, any test of the valuation usefulness of a set of accounting standards is a joint test of the accounting standards and of firms' implementation of those standards.³ Because of the complexity of U.S. GAAP, a non-U.S. firm may interpret and implement U.S. GAAP differently than a U.S. firm. As a result, non-U.S. firms' U.S. GAAP earnings and book values may have different valuation properties than those of U.S. firms.

It is unclear which valuation model will dominate when non-U.S., cross-listed firms report earnings and book values measured under IAS. IAS measurement and recognition standards allow managers more discretion than U.S. GAAP. In particular, IAS allows managers more discretion in the accounting for development costs, property, plant, and equipment (PPE), intangible assets, and interest incurred during the construction of long-term assets. Davis-Friday and Rueschhoff (2001) indicate that these discretionary accounting methods are significant reconciling items for a sample of U.S. foreign registrants that prepare IAS financial statements and U.S. GAAP reconciliations.

Figure 1 summarizes our predictions on how managers' discretion under IAS related to these four accounting methods affects the relative performance of the EC, BV, and RI

² As Palepu et al. (2000) show, $1/r$ implies that a firm's residual income will persist at current levels forever. A lower value implies mean-reversion in residual income. A higher value would imply an exploding profitability process, yielding an infinite price. Thus, $1/r$ is an upper bound.

³ Certain mechanisms in the capital markets act to monitor firms' financial reporting, thereby increasing the likelihood that firms implement standards on a consistent basis. We document that our sample firms have large analyst followings, suggesting that sophisticated information users monitor our sample firms' financial information. We also document that our sample firms are large, multinational, cross-listed firms that are subject to detrimental reputation effects (i.e., higher costs of capital) if they misreport financial information. Firms' auditor choices also provide a form of monitoring. All but one of our sample firms contracts with a Big 6 audit firm. Overall, sample firms appear to be well monitored, suggesting that sample firms implement IAS or U.S. GAAP on a consistent basis.

FIGURE 1
Summary of Accounting Standard Differences and Their Expected Effect on Valuation Models

Standard Related to	IAS Choice Deviating from U.S. GAAP				IAS Choice Aligning with U.S. GAAP			
	Predicted Goodness-of-Fit Effects Relative to U.S. GAAP				Predicted Goodness-of-Fit Effects Relative to U.S. GAAP			
	Treatment	EC Valuation Model	BV Valuation Model	RI Valuation Model	Treatment	EC Valuation Model	BV Valuation Model	RI Valuation Model
Development costs	Capitalize upon feasibility ^a	Increase ^b	Increase ^b	Diffuse ^c	Expense assuming no feasibility ^a	No effect	No effect	No effect
Fixed assets	Upward revalue	Increase ^d	Increase ^d	Decrease ^e	Historical cost	No effect	No effect	No effect
Internally generated intangible assets	Capitalize	Increase ^b	Increase ^b	Diffuse ^c	Expense	No effect	No effect	No effect
Interest costs incurred during construction	Expense	Decrease ^f	Decrease ^f	Diffuse ^c	Capitalize	No effect	No effect	No effect

^a IAS requires managers to capitalize the development costs of a project when certain criteria are met. Management's discretion determines whether and when criteria are met and, consequently, when development costs are capitalized. Therefore, managers have discretion in aligning with or deviating from U.S. GAAP.

^b The accounting method allows management to communicate information about projects' expected future revenue streams, since period costs are capitalized rather than expensed. Capitalization results in the accounting system recognizing a broader scope of economic assets. ^c The accounting choice results in different book values relative to U.S. GAAP, which yields differential residual income valuation implications (i.e., it affects the parameters in the RI valuation model). The overall effect in terms of explanatory power is *ex ante* unclear. ^d The accounting choice results in fair value accounting for fixed assets and the related depreciation charge having greater economic meaning relative to depreciation charges based on historical cost.

^e The accounting choice results in clean surplus violations, i.e., violations of a key RI model assumption.

^f The accounting choice results in less information about projects' expected future revenue streams, since current costs that are expected to generate future revenue streams are expensed rather than capitalized. The balance sheet may not capture all assets and/or assets may be mismeasured.

valuation models. In the first four columns of Figure 1, we assume that IAS managers implement IAS-acceptable methods that deviate from U.S. GAAP. To the extent that managers' discretion under IAS increases the informativeness of IAS firms' earnings and book values, we expect the performance of the EC and BV models to improve relative to the performance of these models when valuing U.S. GAAP accounting variables.

Consider the following example. A manager capitalizes the development costs of an internally developed intangible asset because he expects the asset to generate future revenues. Capitalizing these costs signals the manager's beliefs about the expected future revenue streams resulting from the asset's use. By deferring the recognition of the development costs, present earnings will be larger and will have a higher association with the intangible asset's expected future revenue streams. Because expected future revenues are correlated with the firm's expected future cash flows available to investors, the manager's discretion increases the valuation usefulness of current earnings. In addition, the manager's recognition of the intangible asset yields book values that better reflect the firm's economic assets.

It is not clear *a priori* whether the RI model will perform better when IAS managers use their discretion to implement methods that deviate from U.S. GAAP. Returning to our example, capitalizing costs of an internally developed intangible asset affects the weighting of book values and residual income in the RI model. Because it is impossible to predict how the weights will change in an individual year, we are unable to make a prediction on the relative performance of the RI model.⁴ In addition, it is unclear whether managers' discretion under IAS results in more clean surplus violations than under U.S. GAAP (e.g., voluntary fair value adjustments bypassing the income statement), ultimately impairing the RI model's performance.

Columns (5) through (8) of Figure 1 present our predictions assuming that IAS managers align their accounting measurement and recognition choices with U.S. GAAP. If IAS managers choose to align with U.S. GAAP, then we expect the performance of the EC, BV, and RI models to be similar to that of U.S. GAAP. In summary, Figure 1 highlights our expectations that the relative performance of each valuation model depends on whether a firm chooses IAS accounting measures that deviate or align with U.S. GAAP.

III. SAMPLE SELECTION AND DESCRIPTIVE STATISTICS

We restrict our sample to non-U.S./non-U.K. firms that report 1997 earnings and book values measured in accordance with IAS or U.S. GAAP, and whose shares trade in the SEAQ International Equity Market. SEAQ International is a market for non-U.K. firms that are listed on other exchanges; i.e., it is a market for cross-listed firms. We focus on SEAQ International because it has the largest number of cross-listed firms (Zeff 1998). We restrict our analysis to one market because market regulators are concerned about the valuation implications of allowing dual reporting systems in a common equity market (SEC 2000). Furthermore, differences in listing requirements, monitoring, and microstructures across markets can affect firms' price formation (Frost and Pownall 1994). Thus, drawing a sample from multiple exchanges could impair our ability to draw inferences from our results. We restrict our analysis to 1997 because neither the IAS nor the U.S. GAAP measurement and

⁴ The weights on book value and residual income are of consequence in a single-period implementation of the RI model. In a multiperiod implementation, this particular accounting choice has no effect on the value estimate because the effects of this accounting choice shift the value contributions between book values and future RI without changing the present value of the sum of the value contribution over the life of the asset (see, e.g., Morgan Stanley Dean Witter 1998).

recognition standards changed during that year.⁵ In summary, our sample contains cross-listed, non-U.S./non-U.K. firms that prepare financial information in accordance with an internationally acceptable set of standards, trade in the same market, and depend on foreign capital. In addition, our sample selection criteria identify firms of interest to the SEC in its assessment of the usefulness of IAS for cross-border filings (SEC 2000). These research design choices yield a relatively small sample of 55 firms. We discuss the limitations of our sample in Section V.

The first phase of the sample selection process identifies the sample firms reporting IAS earnings and book values and whose shares are quoted on SEAQ International. The IASC disseminated a list of 191 non-U.S./non-U.K. firms that had claimed at some point to use IAS in their annual reports.⁶ From Bloomberg, we identified 138 firms with foreign listings and sent letters requesting copies of their 1997 annual reports. We needed the annual report to determine whether a firm applies IAS and whether there are any exceptions to its IAS application. We received 106 reports and identified 58 firms that followed IAS in 1997. Of the 58 firms, 26 firms were quoted on SEAQ International three months after their 1997 fiscal year ends.⁷

The second phase of the sample selection process identifies a sample of non-U.S./non-U.K. firms that report U.S. GAAP earnings and book values and that trade their shares on SEAQ International. We identified the intersection of firms that trade on the New York Stock Exchange (NYSE), and thus must file U.S. GAAP financial information with the SEC, and that trade on SEAQ International three months following their 1997 fiscal year-ends. This results in 36 non-U.S., cross-listed firms reporting U.S. GAAP accounting variables.

Panel A of Table 1 presents the distribution of sample firms by domicile. The majority of firms are domiciled in Japan, France, The Netherlands, and Finland. Six IAS firms also report U.S. GAAP accounting variables because they are listed on the NYSE, one each from China, Finland, Germany, and The Netherlands, and two from Sweden. One IAS firm from Sweden voluntarily reports U.S. GAAP information. Thus, seven firms are in both the IAS and U.S. GAAP samples. Panel B of Table 1 reports that the distribution of firms across industries is similar for the U.S. GAAP and IAS samples. Documenting a similar proportion of firms operating in R&D-intensive industries (Chemicals and Pharmaceuticals, Machinery and Computer Hardware, Electrical and Electronics, Transportation Vehicles, and Scientific Instruments; Lev and Sougiannis 1996) is important because of the variation between IAS and U.S. GAAP in accounting for intangible assets. Panel C of Table 1 reports that 22 of the 26 IAS sample firms' shares trade in the U.S., either by a listing with the

⁵ The Comparability Project implemented in 1995 (IASC 1997) and the Core Project completed in 1998 (IASC 1999) changed some of the accounting recognition and measurement rules comprising IAS prior to and after 1997, respectively. In addition, several U.S. GAAP accounting measurement and recognition standard changes occurred in 1995 and 1996. Identifying a period during which the set of IAS and U.S. GAAP measurement and recognition standards are relatively constant increases the likelihood that firms are interpreting and implementing the standards consistently.

⁶ Firms can change their reporting standards (e.g., from IAS to U.S. GAAP), and thus it was important to ascertain whether firms continued to report under IAS.

⁷ In addition to the 26 firms, four Swiss firms used IAS and traded on SEAQ International three months after their 1997 fiscal year-ends. We exclude these four firms from our empirical tests as their market prices per share were more than 10 standard deviations from the mean market price per share of IAS firms, creating a severe scale problem. Including these four firms in our empirical tests does not affect the inferences, although the additional variance due to these firms' exceptionally high share prices inflates the R^2 s of the IAS regressions.

TABLE 1
Non-U.S./Non-U.K. Firms Traded on SEAQ International that Report U.S. GAAP or IAS
Earnings and Book Values

Panel A: By Country^a

<u>Country</u>	<u>U.S. GAAP</u>	<u>IAS</u>
Argentina	1	0
Australia	0	1
Belgium	1	0
China	1	1
Finland	4	3
France	3	4
Germany	3	1
Hong Kong	0	4
Italy	3	1
Japan	7	3
The Netherlands	6	1
Norway	2	2
South Africa	0	1
Spain	2	0
Sweden	3	4
Total	<u>36</u>	<u>26</u>

Panel B: By Industry

<u>Industry (SIC code)</u>	<u>U.S. GAAP</u>	<u>IAS</u>
Mining and Construction (1)	4	0
Manufacturing (2 and 3)	26	19
<i>R&D intensive^b</i>	19	12
<i>Other manufacturing</i>	7	7
Transportation (4)	2	1
Wholesale and Retail (5)	3	4
Finance and Insurance (6)	1	1
Services (7 and 8)	0	1

Panel C: Additional Statistics

	<u>U.S. GAAP</u>	<u>IAS</u>
Traded in the U.S. ^c	36	22
Revaluation of fixed assets ^d	0	14
Capitalization of internally generated intangible assets	0	13

^a U.S. GAAP firms report U.S. GAAP earnings and book values in Form 20-F or in their annual reports. IAS firms report IAS earnings and book values in their annual reports. Seven firms report both U.S. GAAP and IAS accounting variables.

^b R&D-intensive industries are Chemicals and Pharmaceuticals (two-digit SIC code 28), Machinery and Computer Hardware (35), Electrical and Electronics (36), Transportation Vehicles (37), and Scientific Instruments (38).

^c Number of firms traded in the U.S., either on an exchange or in the Pink Sheet market.

^d As disclosed in the footnotes to firms' financial statements.

NYSE ($n = 6$) or by trading in the Pink Sheet market ($n = 16$).⁸ Panel C of Table 1 also reports that 14 IAS firms upward revalue their property, plant, and equipment. Finally, Panel C reports that 13 of 26 IAS sample firms capitalize development costs or internally generated intangible assets. U.S. GAAP requires firms to expense these costs.

Table 2 reports the descriptive statistics for our IAS and U.S. GAAP samples; recall that seven firms are in both. We report statistics including these seven firms, and we conduct our statistical tests with (and without) them.⁹ We collect sample firms' 1997 fiscal-year-end market values, betas, number of exchanges and SEAQ share prices from Bloomberg and analyst following from I/B/E/S. We collect sample firms' net income and book values from their annual reports or SEC filings.

The descriptive statistics indicate some differences between the U.S. GAAP sample firms and the IAS sample firms. First, U.S. GAAP firms' market values are significantly (at the 0.00 level) larger than our IAS firms' market values. Second, the distributions of IAS and U.S. GAAP firms' share prices are significantly (at the 0.05 level) different. Third, U.S. GAAP sample firms' mean market-to-book ratio is larger than that of IAS sample firms; the mean (median) of IAS and U.S. GAAP firms' market-to-book ratios are 2.08 (1.71) and 2.67 (2.25), respectively. There is no statistical difference between our IAS and U.S. GAAP sample firms' betas, analyst following, or exchange listings. In general, there is no statistical difference between our IAS and U.S. GAAP firms' mean net incomes, book values, or residual incomes, where residual income is calculated using a 10 percent discount rate.¹⁰ Although the IAS sample firms are smaller than the U.S. GAAP firms, both samples are composed of large global firms with substantial analyst following.

IV. RESULTS

Primary Results

Panel A of Table 3 reports the results of the IAS valuation analyses.¹¹ We use firms' SEAQ International share prices three months after their fiscal year-ends as the dependent variable.¹² We find that the EC model is the dominant valuation model when firms report under IAS. IAS earnings explain 85 percent of IAS firms' share prices, and the small and insignificant intercept suggests that the EC model is well specified. In the BV model, we find that IAS book values explain 77 percent of price. The Vuong (1989) test reveals that the incremental adjusted R^2 of the EC model relative to the BV is statistically significant

⁸ The Pink Sheet market is a market for over-the-counter stocks. Pink Sheets are the dealer quotations of unlisted stocks published by the National Quotation Bureau. Unlike the National Association of Securities Dealers (NASD) Over-The-Counter Bulletin Board, firms quoted in the Pink Sheet market are not required to file financial information with the SEC. However, non-U.S. firms trading in the Pink Sheet market are required to register with the SEC, using Form F-2 or F-3, if the aggregate world value of their voting stock is the equivalent of at least \$300 million. Neither Form F-2 nor Form F-3 requires firms to report financial information.

⁹ The regression results excluding these seven firms (not tabulated) are similar to the results reported in Table 3 and none of the inferences change.

¹⁰ Prior research finds that the choice of discount rate does not affect cross-sectional analyses of the RI model (Frankel and Lee 1998; Francis et al. 2001). We apply a 10 percent discount rate in our tabulated results. As a sensitivity analysis, we apply discount rates ranging from 6 percent to 16 percent in calculating residual income values. Changing the discount rate does not affect any of our inferences.

¹¹ To test the results' robustness, we repeat all analyses after deleting observations with studentized residuals greater than 2. The results are very similar and all statistical inferences are the same. Furthermore, there are no signs of heteroskedasticity in any of our regressions. As a further sensitivity analysis (not tabulated), we assess the statistical significance of the coefficients using White-adjusted standard errors. Again, the inferences are unchanged.

¹² The use of prices measured three months after firms' fiscal year-ends is subject to the trade-off between accounting information flowing into price and confounding events. As a sensitivity analysis (not tabulated), we repeat all analyses using fiscal year-end prices and draw similar inferences.

TABLE 2
Sample Descriptive Statistics: Non-U.S./Non-U.K. Firms Traded on SEAQ International that Report U.S. GAAP or IAS Earnings and Book Values

Variable ^b	U.S. GAAP Firms					IAS Firms					Statistical tests	
	<i>n</i>	<i>Mean</i>	<i>Median</i>	<i>10th</i>	<i>90th</i>	<i>n</i>	<i>Mean</i>	<i>Median</i>	<i>10th</i>	<i>90th</i>	<i>t-test</i>	<i>Wilcoxon</i>
Market Value	36	9,010	6,288	1,193	20,620	26	4,109	2,452	716	11,617	0.01 (0.00)	0.01 (0.00)
Beta	36	0.88	0.91	0.49	1.20	26	0.98	0.91	0.58	1.66	0.33 (0.26)	0.60 (0.62)
Analyst Following	36	21	21	5	35	26	18	17	4	32	0.29 (0.25)	0.29 (0.26)
Number of Exchanges	36	5	5	4	7	26	5	4.5	3	7	0.35 (0.28)	0.31 (0.32)
Net Income	36	1.79	1.17	0.05	4.84	26	1.14	0.52	0.04	2.54	0.30 (0.32)	0.15 (0.15)
Book Value	36	15.66	8.05	1.70	27.94	26	10.11	4.34	0.79	27.23	0.27 (0.33)	0.10 (0.16)
Price	36	29.81	21.17	1.75	64.34	26	18.04	12.02	0.06	39.04	0.15 (0.17)	0.05 (0.05)
Price-to-Earnings	36	20.85	16.61	7.06	30.89	26	21.52	16.04	10.92	39.46	0.89 (0.84)	0.68 (0.61)
Market-to-Book	36	2.67	2.25	1.10	5.12	26	2.08	1.71	0.70	4.62	0.13 (0.01)	0.12 (0.03)
Residual Income	36	0.40	0.39	-0.63	3.47	26	0.23	0.02	-0.21	1.02	0.77 (0.72)	0.13 (0.09)

^a p-values are based on two-tailed tests; p-values in parentheses exclude the 7 firms that report both U.S. GAAP and IAS variables. The t-test is a two-sample test of difference in means. The Wilcoxon test is the Wilcoxon-Mann-Whitney rank sum test of whether two samples have been drawn from the same distribution.

^b Variable Descriptions: 1997 Fiscal Year-End:

Market Value = price multiplied by the number of shares of common stock outstanding, in millions of pounds (GBP);
 Beta = the market beta for each firm;

Analyst Following = the number of analysts following the firm at its fiscal year-end, per 1/B/E/S;

Number of exchanges = the number of equity markets in which the firm's stock trades;

Net Income = net income per share measured in accordance with the relevant standards (in millions of GBP);

Book Value = book value per share measured in accordance with the relevant standards (in millions of GBP);

Price = price per share at fiscal year end quoted on SEAQ International, converted to pounds (GBP);

Price-to-Earnings = price divided by net income measured in accordance with the relevant standards;

Market-to-Book = price per share divided by book value per share; and

Residual Income = net income per share - 0.10 × (one year lagged book value per share), adjusted for changes in number of shares outstanding.

TABLE 3
Relative Performance of Accounting-Based Valuation Models for Non-U.S./Non-U.K. Firms
Traded on SEAQ International that Report IAS or U.S. GAAP Earnings and Book Values^a

$$\text{Earnings Capitalization Model: } P_i = \alpha_E + \beta_E E_i + \varepsilon_i$$

$$\text{Book Value Model: } P_i = \alpha_{BV} + \beta_{BV} BV_i + v_i$$

$$\text{Residual Income Model: } P_i = \alpha_{RI} + \beta_{RI} BV_i + \gamma_{RI} RI_i + \xi_i$$

Panel A: Valuation of IAS Sample

	<u>Intercept</u>	<u>Earnings</u>	<u>Book Value</u>	<u>Residual Income</u>	<u>Adjusted R²</u>	<u>n</u>
<i>Earnings Capitalization Model</i>						
Coefficient estimate	1.94	16.61			0.85	26
(standard error)	(2.66)	(1.38)				
p-value (=0)	0.47	0.00			0.00	
<i>Book Value Model:</i>						
Coefficient estimate	3.63		1.71		0.77	26
(standard error)	(3.26)		(0.19)			
p-value (=0)	0.28		0.00		0.00	
p-value (=1)			0.00			
<i>Residual Income Model</i>						
Coefficient estimate	2.50		1.39	18.53	0.83	26
(standard error)	(2.81)		(0.19)	(5.86)		
p-value (=0)	0.38		0.00	0.00	0.00	
p-value (=1)			0.05			
Goodness-of-Fit Indicators <i>Model Assumptions Met^b</i> <i>Vuong Test Z Statistic</i> (p-value)						
	<u>Intercept</u>	<u>Coefficient</u>	<u>Adjusted R²</u>	<u>EC</u>	<u>BV</u>	<u>RI</u>
Earnings capitalization	Yes	NA	0.85	—		
Book value	Yes	No	0.77	2.71 (0.01)	—	
Residual income	NA	No	0.83	1.03 (0.30)	2.39 (0.02)	—

Panel B: Valuation of U.S. GAAP Sample

	<u>Intercept</u>	<u>Earnings</u>	<u>Book Value</u>	<u>Residual Income</u>	<u>Adjusted R²</u>	<u>n</u>
<i>Earnings Capitalization Model</i>						
Coefficient estimate	19.73	7.07			0.35	36
(standard error)	(5.91)	(1.60)				
p-value (=0)	0.00	0.00			0.00	
<i>Book Value Model</i>						
Coefficient estimate	10.26		1.41		0.79	36
(standard error)	(3.55)		(0.12)			
p-value (=0)	0.01		0.00		0.00	
p-value (=1)			0.00			
<i>Residual Income Model</i>						
Coefficient estimate	6.62		1.56	3.37	0.87	36
(standard error)	(2.86)		(0.10)	(0.70)		
p-value (=0)	0.03		0.00	0.00	0.00	
p-value (=1)			0.00			

(Continued on next page)

TABLE 3 (Continued)

Goodness-of-Fit Indicators	Model Assumptions Met ^b			Vuong Test Z Statistic (p-value)		
	Intercept	Coefficient	Adjusted R ²	EC	BV	RI
Earnings capitalization	No	NA	0.35	—	—	—
Book value	No	No	0.79	6.71 (0.00)	—	—
Residual income	NA	Yes	0.87	12.92 (0.00)	3.62 (0.00)	—

^a Price is measured three months after fiscal year-end. The valuation attributes are earnings (E), book value (BV), and residual income (RI).

^b The model assumptions are as follows: zero intercept in the EC model and the BV model, BV coefficient equal to 1 in the BV model, and RI coefficient consistent with well-defined linear RI series (below the theoretical upper bound of 10, the inverse of the assumed discount rate of 10 percent).

(at the 0.01 level). We find an insignificant intercept in the BV model, which suggests that the net assets valued by the market are reported on IAS firms' balance sheets; however, the estimated coefficient on book value is greater than 1 (at the 0.00 level), indicating that IAS book values are systematically lower than market values.

The RI model under IAS has an explanatory power of 83 percent, which is not statistically different from that of the EC model. This indicates that adding balance sheet variables to income statement variables does not improve the IAS information set associated with firms' share prices. In addition, the estimated coefficient on residual income (18.53) exceeds the theoretical upper bound of 10 (based on a 10 percent discount rate). Recall that a one-period RI model implementation that links current period accounting variables to price assumes that a firm's residual income follows a linear process. Certain clean surplus violations can affect this property of residual income, thereby violating the assumption that a firm's multiperiod series of residual income can be collapsed into a coefficient times its current year residual income. For example, IAS No. 16, *Property, Plant, and Equipment*, states that the revaluations of fixed assets should be credited to equity unless reversing a previous charge to income (IASC 1997). The "lumpiness" in book value can result in residual incomes that violate the linearity assumption of a one-period RI model, and impair the application of a one-period RI model when firms report under IAS. Fourteen of the 26 IAS firms state in their footnotes that they upward revalue their fixed assets.¹³

Panel B of Table 3 reports the results of the U.S. GAAP valuation analyses.¹⁴ The EC model appears misspecified as we find an intercept that is significantly positive (at the 0.00 level) and the explanatory power of U.S. GAAP earnings for price is only 35 percent.¹⁵ Turning to the BV model, we find that book values explain 79 percent of U.S. GAAP firms' share prices. The coefficient on book value is significantly (at the 0.00 level) greater than 1, suggesting that the market value of firms' reported net assets exceeds the book value of

¹³ Firms' clean surplus violations from prior years have additional measurement effects on the RI model. Without access to time-series data, we cannot empirically disentangle current year from prior year effects. Because we are not reliably able to measure the total effects of firms' clean surplus violations under IAS, we are unable to prepare *pro forma* numbers that would allow us to conduct "as-if U.S. GAAP" analyses.

¹⁴ The results of the analyses include the one U.S. GAAP firm that reports a loss in 1997. Collins et al. (1999) document that the EC valuation model performs differently for firms reporting positive earnings vs. negative earnings. When we exclude from the EC valuation analysis the one U.S. GAAP firm with negative earnings, the R² is 65 percent and the intercept is statistically (at the 0.02 level) different from 0.

¹⁵ These results are similar to those reported for U.S. firms in prior large-sample studies (e.g., Kothari and Zimmerman 1995; Collins et al. 1999).

firms' reported net assets. In addition, the significantly positive (at the 0.01 level) intercept suggests that non-U.S. firms' U.S. GAAP reporting results in off-balance sheet net assets.

The RI model is the dominant valuation model for our U.S. GAAP sample. The RI model explains 87 percent of U.S. GAAP firms' share prices. The incremental adjusted R^2 of the RI model relative to the EC and the BV models, respectively, is statistically significant (both at the 0.00 level). In addition, the coefficient on residual income (3.37) is below the theoretical upper bound, consistent with a mean-reverting linear residual income process. The positive intercept in the RI model is consistent with the claim that U.S. GAAP commonly results in conservative accounting for firms' assets.

Having documented the dominant valuation model within each reporting regime, we statistically test for coefficient differences between IAS and U.S. GAAP by conducting a categorical analysis in which we pool the IAS and U.S. GAAP samples and re-estimate the three valuation models using dummy variables to distinguish between IAS and U.S. GAAP firms. The results (untabulated) confirm that the differences in the estimated parameters are statistically significant. Significant differences in the parameters indicate that the summary accounting variables under IAS and U.S. GAAP have differential valuation properties, consistent with equity market participants' and market regulators' concerns that IAS and U.S. GAAP accounting variables may not be comparable inputs to investment decisions. Although it is not possible to disentangle the mechanical effects of IAS vs. U.S. GAAP accounting standards from effects caused by differences in characteristics of firms that choose to report under IAS vs. U.S. GAAP, the results of the categorical analysis suggest that the relations between prices and accounting numbers differ between IAS and U.S. GAAP firms.

U.S. GAAP Reporting-Choice Analysis

Currently, non-U.S. firms reporting U.S. GAAP accounting variables can choose either to prepare U.S. GAAP financial reports or to prepare reconciliations from domestic (or IAS) earnings and book values to U.S. GAAP earnings and book values. We explore whether this reporting choice influences the relative performance of the EC, BV, and RI models. We partition the U.S. GAAP sample firms into two groups: firms that prepare U.S. GAAP reconciliations (U.S. GAAP reconcilers) and firms that prepare U.S. GAAP financial statements (U.S. GAAP reporters).¹⁶

Table 4 presents the relative valuation analyses for the 17 U.S. GAAP reconcilers and the 12 U.S. GAAP reporters. We find that the RI model is the dominant valuation model, followed by the BV and then the EC model, regardless of U.S. GAAP reporting method. The model rankings are consistent with our main U.S. GAAP results. Based on the explanatory power of the models, however, the accounting information of U.S. GAAP reconcilers appears less useful in valuation than that of U.S. GAAP reporters. Specifically, the explanatory powers of the RI, the BV, and the EC models for reporters are 98, 77, and 43 percent, respectively, vs. 74, 60, and 9 percent, respectively, for the reconcilers. The negative coefficient on residual income for reconcilers is not a direct violation of model assumptions. It does, however, imply an oscillating pattern of residual income, which is economically nonintuitive. The periodic adjustment from domestic GAAP to U.S. GAAP in reconciliations may drive this result. The lower explanatory power of the valuation models for U.S. GAAP reconcilers relative to U.S. GAAP reporters is consistent with anecdotal

¹⁶ The reconcilers (17 firms) are domiciled in 10 countries, whereas the reporters (12 firms) are from 5 countries, with 3 countries being represented in the intersection of the two groups. This analysis excludes the seven firms that report both IAS and U.S. GAAP variables.

TABLE 4
U.S. GAAP Reporting-Choice Analysis

Comparison of firms that prepare U.S. GAAP reconciliations to firms that prepare U.S. GAAP financial statements

Earnings Capitalization Model: $P_i = \alpha_E + \beta_E E_i + \varepsilon_i$

Book Value Model: $P_i = \alpha_{BV} + \beta_{BV} BV_i + v_i$

Residual Income Model: $P_i = \alpha_{RI} + \beta_{RI} BV_i + \gamma_{RI} RI_i + \xi_i$

Panel A: Valuation of U.S. GAAP Reconcilers

	<u>Intercept</u>	<u>Earnings</u>	<u>Book Value</u>	<u>Residual Income</u>	<u>Adjusted R²</u>	<u>n</u>
<i>Earnings Capitalization Model</i>						
Coefficient estimate	17.22	2.92			0.09	17
(standard error)	(5.62)	(1.79)				
p-value (= 0)	0.01	0.12			0.12	
<i>Book Value Model</i>						
Coefficient estimate	5.32		1.79		0.60	17
(standard error)	(4.60)		(0.36)			
p-value (=0)	0.27		0.00		0.00	
p-value (=1)			0.04			
<i>Residual Income Model</i>						
Coefficient estimate	3.12		2.51	-4.43	0.74	17
(standard error)	(3.74)		(0.37)	(1.43)		
p-value (=0)	0.42		0.00	0.01	0.00	
p-value (=1)			0.00			

Panel B: Valuation of U.S. GAAP Reporters

	<u>Intercept</u>	<u>Earnings</u>	<u>Book Value</u>	<u>Residual Income</u>	<u>Adjusted R²</u>	<u>n</u>
<i>Earnings Capitalization Model</i>						
Coefficient estimate	35.78	8.25			0.43	12
(standard error)	(13.81)	(2.69)				
p-value (=0)	0.30	0.01			0.01	
<i>Book Value Model</i>						
Coefficient estimate	11.65		1.38		0.77	12
(standard error)	(10.64)		(0.23)			
p-value (=0)	0.30		0.00		0.00	
p-value (=1)			0.13			
<i>Residual Income Model</i>						
Coefficient estimate	6.38		1.65	5.00	0.98	12
(standard error)	(3.25)		(0.07)	(0.50)		
p-value (=0)	0.08		0.00	0.00	0.00	
p-value (=1)			0.00			

evidence that the complexity of U.S. GAAP leads non-U.S. firms to make material measurement errors when they apply U.S. GAAP only on a periodic basis (Bradley 2001).

Within-Firm Analysis

Prior research investigating the comparability of IAS to U.S. GAAP uses non-U.S. firms that are required to file U.S. GAAP accounting data with the SEC (Harris and Muller 1999). However, Pownall and Schipper (1999) suggest that IAS firms required to prepare U.S. GAAP financial information may choose alternatives available under IAS that align with U.S. GAAP rather than choosing IAS-acceptable methods that maximize their financial-reporting objectives. We explore this conjecture by estimating the valuation models using the IAS accounting variables and then using the U.S. GAAP accounting variables of the seven sample firms that report under IAS and reconcile to U.S. GAAP. Recall from Figure 1 that if an IAS firm implements the accounting treatment that aligns with U.S. GAAP, we predict no differences in the EC, BV, and RI models' performances relative to U.S. GAAP.

Table 5 presents the EC, BV, and RI analyses using the seven IAS firms that also report U.S. GAAP accounting variables, i.e., the constrained IAS firms.¹⁷ The valuation differences between IAS and U.S. GAAP appear less pronounced for these constrained IAS firms than for the IAS and U.S. GAAP analyses reported in Table 3. The financial statement footnotes of our IAS firms reveal that only one of the seven constrained IAS firms voluntarily upward revalues assets, whereas 13 of the 19 unconstrained IAS firms upward revalue assets. Only one of the seven constrained IAS firms capitalizes development costs or internally generated intangible assets, compared to 12 of the 19 unconstrained IAS firms. We are unable to quantify firms' methods of accounting for interest incurred during long-term construction projects because firms did not disclose in their summaries of accounting policies their methods of accounting for interest incurred during construction. Although the small sample size limits our ability to draw definitive conclusions, these findings are consistent with the conjecture that IAS firms that list in the U.S. are more likely to choose methods that align with U.S. GAAP rather than choosing the IAS treatments that maximize the reporting benefits allowable under IAS. Thus, using non-U.S. firms' Form 20-F reconciliations as a basis to determine the valuation usefulness of IAS relative to U.S. GAAP may underestimate the valuation implication differences between the two alternative sets of internationally acceptable accounting standards.

Scale Sensitivity Analysis

Our tests measure intrinsic value as market price per share. We posit that we capture any scale differences between our IAS and U.S. GAAP firms by measuring the accounting variables on a per-share basis. Easton (1998), however, argues that price per share may not be a sufficient scale control in tests of the relation between price and accounting variables. He suggests scaling by book value of equity. However, including an accounting variable as a scaler would confound our comparisons of the valuation properties of IAS and U.S. GAAP accounting variables. Barth and Kallapur (1996) argue that it is econometrically preferable to include an independent scaler variable rather than to divide all variables by the scaler to control for scale differences.

¹⁷ For the six IAS firms that report U.S. GAAP information to the SEC, IAS income (book value) is, on average, 2.3 percent (1.79 percent) lower than U.S. GAAP income (book value). These differences are similar to those reported in prior studies that examine the Form 20-F reconciliations of IAS firms (Harris and Muller 1999; Davis-Friday and Rueschhoff 2001). The differences between IAS and U.S. GAAP accounting variables for the one firm voluntarily applying U.S. GAAP are as follows: IAS income (book value) is 7.3 percent (5.3 percent) lower than U.S. GAAP income (book value).

TABLE 5
Within-Firm Analysis

Firms that use IAS for financial reporting but also provide U.S. GAAP reconciliations.^a

Earnings Capitalization Model: $P_i = \alpha_E + \beta_E E_i + \varepsilon_i$

Book Value Model: $P_i = \alpha_{BV} + \beta_{BV} BV_i + v_i$

Residual Income Model: $P_i = \alpha_{RI} + \beta_{RI} BV_i + \gamma_{RI} RI_i + \xi_i$

Panel A: Earnings Capitalization Model^b

	<u>Intercept</u>	<u>Earnings</u>	<u>Adjusted R²</u>	<u>n</u>
<i>IAS</i>				
Coefficient estimate	3.39	18.45	0.76	7
(standard error)	(4.00)	(4.11)		
p-value (=0)	0.44	0.01	0.01	
<i>U.S. GAAP</i>				
Coefficient estimate	7.77	14.20	0.46	7
(standard error)	(5.39)	(5.70)		
p-value (=0)	0.21	0.06	0.06	

Panel B: Book Value Model

	<u>Intercept</u>	<u>Book Value</u>	<u>Adjusted R²</u>	<u>n</u>
<i>IAS</i>				
Coefficient estimate	8.22	2.48	0.26	7
(standard error)	(6.97)	(1.40)		
p-value (=0)	0.29	0.14	0.14	
p-value (=1)		0.34		
<i>U.S. GAAP</i>				
Coefficient estimate	8.29	2.42	0.25	7
(standard error)	(7.10)	(1.40)		
p-value (=0)	0.30	0.15	0.15	
p-value (=1)		0.36		

Panel C: Residual Income Model

	<u>Intercept</u>	<u>Book Value</u>	<u>Residual Income</u>	<u>Adjusted R²</u>	<u>n</u>
<i>IAS</i>					
Coefficient estimate	1.08	2.51	14.67	0.79	7
(standard error)	(4.18)	(0.75)	(3.97)		
p-value (=0)	0.81	0.03	0.02	0.02	
p-value (=1)		0.11			
<i>U.S. GAAP</i>					
Coefficient estimate	0.29	3.16	11.81	0.72	7
(standard error)	(5.06)	(0.89)	(3.85)		
p-value (=0)	0.96	0.02	0.04	0.04	
p-value (=1)		0.07			

^a Six of the seven firms file Form 20-F; one firm voluntarily provides U.S. GAAP numbers.

^b None of the sample firms report losses under IAS or U.S. GAAP.

We assess potential scale issues in two ways. First, we use the number of shares outstanding at fiscal year-end as an independent scale variable. Second, we use the number of employees at fiscal year end (as reported in *Worldscope*) as an independent scale variable.¹⁸ The inferences drawn from our original analyses are unchanged.

V. DISCUSSION

Primary Findings

We use a sample of non-U.S./non-U.K. firms that trade on London's SEAQ International to explore the valuation properties of cross-listed firms' IAS and U.S. GAAP earnings and book values. We draw inferences on the valuation properties of our sample firms' IAS and U.S. GAAP accounting variables based on the relative performance of the earnings-capitalization (EC), book value (BV), and residual income (RI) valuation models. We find that the BV model performs reasonably well under either set of standards, although both sets of standards produce conservative measurements of net assets. Our results indicate that the RI model is the dominant valuation model when non-U.S., cross-listed firms report under U.S. GAAP. Since our U.S. GAAP firms also trade on the NYSE, this result suggests that U.S. foreign registrants' implementation of U.S. GAAP produces earnings and book values that have similar valuation properties to those of U.S. domestic registrants.

We find that the EC model is the dominant valuation model when non-U.S., cross-listed firms report under IAS. The EC model also appears well specified for the seven firms that report under IAS and reconcile to U.S. GAAP; however, the RI model has higher explanatory power for these IAS firms. Thus, the valuation properties of the IAS earnings and book values reported by these seven firms appear more similar to U.S. GAAP earnings and book values than do those of IAS firms that do not reconcile to U.S. GAAP. This finding, if confirmed in a large sample study, would suggest that policymakers should not judge the acceptability of IAS based only on the examination of firms that use IAS and reconcile to U.S. GAAP, because such firms' financial statements are likely to have quite different valuation properties relative to other IAS users' financial statements.

Generalizability

Restricting our sample to one equity market, SEAQ International, controls for market institutional features that could otherwise affect price formation and the measured relations between share prices and accounting variables. However, equity markets other than SEAQ International allow foreign registrants to file financial information prepared in accordance with IAS or U.S. GAAP (e.g., the Frankfurt Exchange). It is unclear if our inferences will generalize to less developed equity markets or to a sample of firms pooled across multiple equity markets.

We restrict our analysis to 1997 because the set of measurement standards comprising IAS and U.S. GAAP were relatively constant at that time. It is not clear whether the IAS results will continue to hold after the IASC's Core Project is fully implemented in 2001. Furthermore, it remains an open question whether our IAS results hold for IAS firms that are not cross-listed. Firms in some developing countries use IAS as their domestic standards

¹⁸ Specifically, we use the following regressions:

$$\begin{aligned} \text{(ScaleEC)} \quad \text{MktCap}_i &= \alpha_E + \beta_E E_i + \delta_E \text{SCALE}_i + \varepsilon_i \\ \text{(ScaleBV)} \quad \text{MktCap}_i &= \alpha_{BV} + \beta_{BV} BV_i + \delta_{BV} \text{SCALE}_i + \nu_i \\ \text{(ScaleRI)} \quad \text{MktCap}_i &= \alpha_{RI} + \beta_{RI} BV_i + \gamma_{RI} RI_i + \delta_{RI} \text{SCALE}_i + \xi_i \end{aligned}$$

where MktCap_i is the market capitalization of firm i , E_i is firm i 's total earnings, BV_i is firm i 's book value of equity, RI_i is firm i 's residual income, and SCALE_i is the alternative measure of scale.

(IASC 2000) and may not be subject to the same monitoring as our sample of cross-listed IAS firms.

We also explore the valuation implications of allowing U.S. foreign registrants the reporting choice of preparing a U.S. GAAP reconciliation or U.S. GAAP financial statements. The lower explanatory power of the valuation models for U.S. GAAP reconcilers relative to U.S. GAAP reporters suggests that the extra disclosures inherent in U.S. financial statements enhance the valuation properties of earnings and book values. Future research can investigate whether the results of our U.S. GAAP reconciliation vs. U.S. GAAP reporting analysis generalize to the entire population of U.S. foreign registrants.

REFERENCES

- Ashbaugh, H. 2001. Non-U.S. firms' accounting standard choices. *Journal of Accounting and Public Policy* 20: 129–153.
- Ball, R., S. P. Kothari, and A. Robins. 2000. The effect of international institutional factors on properties of accounting earnings. *Journal of Accounting and Economics* 29: 1–51.
- Barker, R. G. 1998. The role of dividends in valuation models used by analysts and fund managers. Working paper, Cambridge University, U.K.
- Barth, M. E., and S. Kallapur. 1996. The effects of cross-sectional scale differences on regression results in empirical accounting research. *Contemporary Accounting Research* 13: 527–567.
- Bradley, L. 2001. KPMG current issues impacting practice. Presentation at the American Accounting Association International Section Midyear Meeting, April 5–7, 2001, Phoenix, Arizona.
- Collins, D., M. Pincus, and H. Xie. 1999. Equity valuation and negative earnings: The role of book value of equity. *The Accounting Review* 74: 29–61.
- Davis-Friday, P., and N. Rueschhoff. 2001. An analysis of the implications of the IASC comparability project for IAS adopters in the U.S. *Asia Pacific Journal of Accounting and Economics* 8.
- Easton, P. D., and T. S. Harris. 1991. Earnings as an explanatory variable for returns. *Journal of Accounting Research* 29: 19–36.
- . 1998. Discussion of revalued financial, tangible, and intangible assets: Association with share prices and non-market-based value estimates. *Journal of Accounting Research* 36: 235–247.
- Financial Accounting Standards Board (FASB). 1978. *Objectives of Financial Reporting by Business Enterprises*. Statement of Financial Accounting Concepts No. 1. Stamford, CT: FASB.
- . 1999. *The IASC-U.S. Comparison Project: A Report on the Similarities and Differences between IASC Standards and U.S. GAAP*. 2nd edition. Stamford, CT: FASB.
- Francis, J., and K. Schipper. 1999. Have financial statements lost their relevance? *Journal of Accounting Research* 37: 319–352.
- , P. Olsson, and D. Oswald. 2000. Comparing the accuracy and explainability of dividend, free cash flow and abnormal earnings equity value estimates. *Journal of Accounting Research* 38: 45–70.
- , ———, and ———. 2001. Using mechanical earnings and residual income forecasts in equity valuation. Working paper, Duke University and London Business School.
- Frankel, R., and C. Lee. 1998. Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics* 25: 283–319.
- Frost, C., and G. Pownall. 1994. A comparison of the stock price response to earnings disclosures in the United States and the United Kingdom. *Contemporary Accounting Research* 11: 59–84.
- , and W. Kinney. 1996. Disclosure choices of foreign registrants in the United States. *Journal of Accounting Research* 34: 67–84.
- Goldman Sachs. 1999. Accounting/portfolio strategy. *Goldman Sachs Investment Research* (May 28): 1–40.
- Harris, M., and K. Muller. 1999. The market valuation of IAS versus U.S. GAAP accounting measures using Form 20-F reconciliations. *Journal of Accounting and Economics* 26: 285–312.
- Holthausen, R. W., and R. L. Watts. 2000. The relevance of the value relevance literature for financial accounting standard setting. Working paper, University of Pennsylvania and University of Rochester.

- International Accounting Standards Committee (IASC). 1997. *International Accounting Standards*. London, U.K.: IASC.
- . 1999. *International Accounting Standards*. London, U.K.: IASC.
- . 2000. Use of IAS in countries around the world. March 3. Available at: <http://www.iasc.org.uk/cmt/1001.asp>.
- Kothari, S. P., and J. L. Zimmerman. 1995. Price and return models. *Journal of Accounting and Economics* 20: 155–192.
- Lev, B., and T. Sougiannis. 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics* 21: 107–138.
- Morgan Stanley Dean Witter. 1998. *Apples to Apples—Accounting for Value in World Markets*. New York, NY: Morgan Stanley Dean Witter International Investment Research.
- Myers, J. N. 1999. Implementing residual income valuation with linear information dynamics. *The Accounting Review* 74: 1–28.
- New York Stock Exchange (NYSE). 1999. September 9. Available at: <http://www.nyse.com/international/international.html>.
- Nissim, D., and S. H. Penman. 1999. Ratio analysis and equity valuation. Working paper, Columbia University.
- Ohlson, J. 1995. Earnings, book values and dividends in equity valuation. *Contemporary Accounting Research* 11: 661–687.
- Palepu, K., P. Healy, and V. Bernard. 2000. *Business Analysis and Valuation: Using Financial Statements*. 2nd edition. Cincinnati, OH: South-Western College Publishing.
- Penman, S., and T. Sougiannis. 1998. A comparison of dividend, cash flow, and earnings approaches to equity valuation. *Contemporary Accounting Research* 15: 343–384.
- Pownall, G., and K. Schipper. 1999. Implications of accounting research for the SEC's consideration of International Accounting Standards for U.S. securities offerings. *Accounting Horizons* 13: 259–280.
- Securities and Exchange Commission (SEC). 2000. *International Accounting Standards*. February 16, 2000. Release Nos. 33-7801, 34-42430; International Series Release No. 1215; File No. S7-04-00. Washington, D.C.: Government Printing Office.
- Vuong, Q. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307–333.
- Zeff, S. 1998. The IASC's core standards: What will the SEC do? *Journal of Financial Statement Analysis* 4: 67–78.

The Interaction between Accrual Management and Hedging: Evidence from Oil and Gas Firms

Morton Pincus

The University of Iowa

Shivaram Rajgopal

University of Washington

ABSTRACT: This research investigates whether oil and gas producing firms use abnormal accruals and hedging with derivatives as substitutes to manage earnings volatility. Firms engaged in oil exploration and drilling are exposed to two kinds of risks that can cause earnings volatility: oil price risk and exploration risk. Firms can use abnormal accrual choices and/or derivatives to reduce earnings volatility caused by oil price risk, but cannot directly hedge the operational risk of unsuccessful drilling. Because hedging and using abnormal accruals are costly activities, and because prior research suggests managers do not eliminate all volatility (Haushalter 2000; Barton 2001), we expect that, at the margin, managers will use these smoothing mechanisms as substitutes to manage earnings volatility. Our results suggest a sequential process whereby managers of oil and gas producing firms first determine the extent to which they will use derivatives to hedge oil price risk, and then, especially in the fourth quarter, manage residual earnings volatility by trading off abnormal accruals and hedging with derivatives to smooth income.

Keywords: *hedging; derivatives; income smoothing; abnormal or discretionary accruals; oil and gas firms.*

Data Availability: *All data used in this research are from publicly available sources.*

We appreciate the comments and suggestions we received from the anonymous referees, Mary Barth (the associate editor), workshop participants at the joint Oregon-Washington-UBC workshop, the University of Pittsburgh, the University of Wisconsin, and the 1999 American Accounting Association Annual Meeting, and from Lisa Bryant, Dave Burgstahler, Neil Fargher, Gerry Feltham, Steve Fortin, Jim Jiambalvo, Mark Kohlbeck, Eric Noreen, Joe Paperman, Terry Shevlin, and D. Shores. Professor Pincus gratefully acknowledges the financial support of the Tippie College of Business at The University of Iowa.

*Submitted July 1999
Accepted August 2001*

I. INTRODUCTION

An active stream of research investigates whether managers smooth income by taking actions to reduce the time-series variability in reported earnings (e.g., Ronen and Sadan 1981; Schipper 1989; Hunt et al. 1996). Schrand and Elliott (1998, 276) note that managers frequently cite the objective of controlling accounting risk, "the risk associated with variability in accounting amounts." DeFond and Park (1997) provide evidence that managers smooth income because of job security concerns; research on bond default risk (Smith and Stulz 1985; Trueman and Titman 1988), income taxes (Graham and Smith 1999), and information asymmetry (DeMarzo and Duffie 1995) demonstrates that reducing earnings volatility can benefit shareholders. Barth et al. (1999) document higher price-earnings multiples for firms with steadily increasing earnings, and a decline in price-earnings multiples when earnings fall after a period of increasing earnings. Yet the process by which managers smooth earnings is not well understood. To expand our understanding of this process, we examine the relation between two alternative mechanisms that managers of oil and gas producing firms can use to manage earnings variability: abnormal accruals and hedging with derivatives. Hedging dampens volatility by directly affecting the distribution of underlying cash flows, whereas smoothing with abnormal accruals directly affects only earnings volatility. In this study we ask whether, at the margin, managers of oil and gas producing firms treat hedging and accrual management as substitute mechanisms for smoothing.

This question is important for several reasons. Lambert (1984) argues that firms have incentives to use both accounting choices and real actions to smooth income. Prior research has considered real activities (e.g., Hand 1989) or accounting decisions (e.g., DeFond and Park 1997) that smooth income, but (with a few exceptions, e.g., Barton [2001]) has generally not considered both. Our study investigates whether (and how) managers draw from a portfolio of accounting tools (accruals and the full cost/successful efforts methods choice) and economic tools (e.g., hedging and diversification of operations) to manage earnings volatility in ways that reflect differences in incentives and in the costs and benefits of using the tools. Hence, researchers seeking to explain risk management behavior with regard to earnings volatility should find our evidence on the interaction between managers' operating actions and accounting decisions relevant to their research. Managers can use economic tools such as hedging along with accounting tools such as abnormal accruals to smooth income, and it is as important to study why managers choose a particular method of income smoothing as it is to understand why they smooth income in the first place. Similarly, regulators investigating income-smoothing activities (e.g., Loomis 1999) and standard setters considering accounting rules that constrain managers' accounting choices and inhibit their ability to smooth income should find our research informative because it allows for the possibility that managers may substitute between accounting and economic tools to smooth income as the costs and effectiveness of one tool change relative to that of alternative tools.

We focus on oil and gas firms that are primarily engaged in exploration and drilling. Two types of industry-specific risks affect the volatility of their earnings and, thus, their accounting risk. The first risk, fluctuations in oil prices, is due to market factors that are beyond management's control. The second risk arises from the firm's drilling success (Malmquist 1990; Fargher et al. 1997). Oil and gas producers can use derivatives to hedge oil price risk, but not the risk of unsuccessful exploration. There are no markets comparable to oil futures markets in which a firm can hedge its oil exploration risk.

If managers always preferred less volatility, then they would use all available techniques to reduce earnings volatility. However, Haushalter (2000) documents that oil and gas producers do not hedge all of their exposure to oil price risk, and Barton (2001) argues that, in general, managers strive for some nonzero level of earnings volatility. Managers can have incentives not to hedge if, for example, either they hold stock options and the value of their options increases in volatility (Tufano 1996), or they seek to coordinate risk management strategies and perhaps expose their firms to core-activity-risks that are associated with higher expected returns (Schrand and Unal 1998). Moreover, efficient hedging requires expertise, and oil and gas producers face basis risk—the risk that changes in the value of derivatives that are available for hedging purposes are not highly correlated with changes in the value of the firm's specific oil and gas production from the particular locations that the firm wishes to hedge. On the other hand, it is probably less costly to obtain expertise to manage accruals; accrual entries immediately affect reported earnings, and managers can alter accrual decisions after year-end. Of course, generally accepted accounting principles (GAAP) and scrutiny of firms by external parties such as independent auditors constrain managers' accrual choices, thereby impeding their ability to use abnormal accruals to smooth income. Because both hedging and smoothing with abnormal accruals are costly and imperfect mechanisms for managing earnings volatility, and because prior research suggests that managers do not eliminate all volatility, we expect managers to trade off one smoothing tool for the other at the margin.

Our investigation of the way oil and gas firms use hedging and smoothing with abnormal accruals to manage earnings volatility has several key features. First, to allow for the substantive differences between firms that hedge and those that do not (Geczy et al. 1997; Haushalter 2000), we analyze separately the following decisions: (1) whether to hedge, and (2) if hedged, the amount of hedging (Cragg 1971; Schrand 1994; Haushalter 2000; Barton 2001). Second, hedging and smoothing with abnormal accruals likely are endogenous elements of a firm's overall risk management strategy. We use a simultaneous equations system in which the regression explaining the extent of hedging includes the empirical proxy for the extent of smoothing with abnormal accruals, and the regression explaining the extent of smoothing with abnormal accruals includes the empirical proxy for the extent of hedging. Third, it is the abnormal component of accruals that is relevant to our investigation. Accordingly, we disaggregate total accruals and estimate the "normal" and "abnormal" components. Fourth, oil and gas producers also decide whether to use the full cost or successful efforts method to account for exploration costs, so we control for a firm's choice of the full cost or successful efforts method, and also incorporate this choice into our estimation of abnormal accruals. Fifth, we control for other determinants of hedging and smoothing with abnormal accruals.

Two prior studies have examined the relation between accounting choice and hedging. Petersen and Thiagarajan (2000) report case study evidence of two gold-mining firms; one managed risk with derivatives while the other used accounting estimates to smooth earnings. Barton (2001) documents a simultaneous and negative relation between foreign exchange and interest rate derivative holdings and abnormal accruals in a broad cross-sectional subset of *Fortune 500* firms. His results are largely consistent with our substitution hypothesis.

Our study complements and extends Barton (2001) by focusing on commodity derivatives in a single industry. This allows us to identify and measure more precisely our sample firms' inherent market and operational risks (Hughes 2000). Thus, we can compute a hedging ratio that more accurately pinpoints the proportion of risk exposure hedged, and we can estimate an important operational risk (exploration risk) and examine its interaction with hedging and smoothing with abnormal accruals. Focusing on a single industry also

allows us to hold production functions relatively constant in the cross-sectional analysis and to identify the effect of an important accounting choice (full cost or successful efforts) on abnormal accruals. In contrast to Barton's (2001) use of the absolute value of abnormal accruals, we compute a smoothing ratio—the standard deviation of earnings before abnormal accruals to the standard deviation of reported earnings—that captures the direct effect of using abnormal accruals to smooth income. Thus, we re-examine the substitution hypothesis regarding the extent of hedging and smoothing with abnormal accruals in a single industry setting where our endogenous variables are likely less subject to measurement error. The main costs of our single-industry design are a smaller sample and an inability to generalize beyond oil and gas producing firms.

Our results indicate that the extent of smoothing with abnormal accruals is not a significant determinant of the amount of hedging. In contrast, the extent of hedging is a significant determinant of the extent of smoothing with abnormal accruals. Specifically, we find that even after controlling for factors affecting cross-sectional differences in incentives to smooth, the more managers hedge with derivatives, the less they smooth with abnormal accruals. The results are consistent with a sequential process whereby managers first make hedging decisions, and then, at the margin, substitute between abnormal accruals and hedging with derivatives to smooth earnings. Detailed analysis indicates that fourth-quarter abnormal accruals reflect this trade-off between the two smoothing mechanisms.

Our inference of a sequential hedging-then-abnormal-accruals decision process contrasts with Barton's (2001) overall conclusion of a simultaneous process whereby abnormal accruals affect hedging, as well as hedging affecting abnormal accruals. We believe that the difference in our inferences is most likely due to our ability to measure more accurately the extent of hedging and smoothing with abnormal accruals, or to unique features of the oil and gas industry. Unfortunately, however, we cannot rule out the possibility that lower power tests resulting from our smaller sample size contributed to our conclusion that abnormal accruals do not play a significant role in explaining the extent of hedging.

We organize the remainder of the paper as follows. First, in Section II we develop the motivation for expecting managers of oil and gas producers to view hedging oil price risk and smoothing with abnormal accruals as substitute devices for managing earnings volatility. We then discuss the empirical design in Section III and identify the explanatory variables for the hedging and the smoothing with abnormal accruals regressions. Section IV presents descriptive statistics, the primary results, and additional analyses, and Section V concludes.

II. RESEARCH HYPOTHESIS

Exploration (or drilling) risk is the risk that exploring for oil and gas will result in "dry" wells. Exploration risk introduces variation in the quantities of oil and gas produced, thereby inducing variability in a firm's cash flows. Cash flow (and thus earnings) volatility also arises from oil price risk—the risk of fluctuating revenues due to volatile oil and gas prices. A firm can reduce oil price risk by hedging with derivative instruments; however, such instruments cannot hedge exploration risk.

If managers preferred to minimize cash flow volatility and earnings volatility, then oil and gas producing firms would hedge all oil price risk they face *and* use other mechanisms to reduce the volatility induced by unhedgeable exploration risk. That is, managers would use these volatility-controlling mechanisms in a complementary, or reinforcing, fashion to reduce volatility. Hence, in addition to hedging oil price risk to reduce cash flow volatility, managers would use abnormal accruals (AACs) to smooth reported earnings—for example,

to maximize share price (e.g., Barth et al. 1999), to lower the firm's expected tax liability (Smith and Stulz 1985), or perhaps to communicate private information to investors about expected future cash flow volatility (Subramanyam 1996). However, managers can have incentives against minimizing volatility. For instance, managers with stock options may opt not to hedge in the hope of increasing stock price volatility (Tufano 1996; Rajgopal and Shevlin 2002). Further, managers may want to increase their exposure to particular risks where they anticipate higher returns, especially in core activities, while at the same time hedging volatility from other risks. Schrand and Unal (1998) find such evidence in the thrift industry and conclude it is indicative of a coordinated risk-management strategy. In our setting, this suggests that managers might hedge oil price risk while exposing their firms to exploration risk, or perhaps hedge oil price risk less, the lower their firm's exploration risk.

Moreover, hedging and smoothing with AACs are costly and imperfect tools for managing volatility, and thus at some point one or the other may not be cost effective. Prior research on hedging (Mian 1996; Geczy et al. 1997; Haushalter 2000) links firm size and hedging. Larger firms have the economies of scale in information and transactions costs to hedge efficiently (e.g., hiring personnel with the experience to manage a derivatives program). In addition, value changes in oil and gas produced in a firm's locations may not necessarily be highly correlated with the value changes in the oil and gas produced in other locations that are used as the benchmarks for derivatives; this gives rise to basis risk (Haushalter 2000). Further, it is likely more costly to obtain expertise in hedging than expertise in accrual management, and managers can make important current-period accrual decisions after year-end. Of course, there are impediments to using AACs to smooth. These include monitoring by auditors and financial analysts, accrual reversals, and other constraints under GAAP.

In our sample, oil price risk and exploration risk are positively correlated ($\rho = 0.16$, $p = 0.07$, two-tailed test).¹ Thus, managers can use both hedging and smoothing with AACs to reduce earnings volatility induced by these risks. We therefore expect managers to use both mechanisms to manage volatility, consistent with Barton's (2001) evidence. However, the costs and limitations of both hedging and using AACs as tools for managing volatility, as well as differing incentives regarding the management of volatility, suggest that managers do not use hedging with derivatives and smoothing with AACs to *eliminate* volatility. Instead, we expect that once managers decide to use both hedging and smoothing with AACs to smooth earnings, they will make trade-offs between these two smoothing instruments at the margin to achieve some benchmark, nonzero level of volatility.

Our basic research hypothesis is as follows:

H1: *Ceteris paribus*, managers of oil and gas firms use hedging with derivatives and smoothing with abnormal accruals as substitute mechanisms at the margin to manage earnings volatility induced by oil price risk and exploration risk.

As discussed in the next section, we employ a simultaneous equation design and control for factors affecting cross-sectional differences in incentives for smoothing.

¹ We estimate an oil price beta for each firm-year as our measure of oil price risk, by regressing firms' daily stock returns on market returns and percentage changes in oil prices. The mean oil price beta is reliably positive, consistent with firms not being fully hedged. Note that if oil price and exploration risks were negatively related, hedging oil price risk would *increase* earnings volatility, because earnings shocks from oil price fluctuations would be hedged and therefore would not be available to offset shocks from exploration activities.

III. EMPIRICAL DESIGN

Empirical Models

Prior research (Geczy et al. 1997, 1999) reports significant differences between hedgers and non-hedgers, and Haushalter (2000) and Barton (2001) find the determinants of the decision whether to hedge differ from the determinants of the extent of hedging, given that a firm hedges. Thus, our analysis separates the decision of whether to hedge from the decision of how much to hedge (Schrand 1994; Haushalter 2000). Then, given that a firm hedges, we allow managers' decisions about the extent of hedging and smoothing with AACs to be simultaneous; i.e., these decisions are endogenous to an entity's risk management strategy, and decisions about each can affect the other (Barton 2001). Therefore, (1) the extent of smoothing with AACs is an endogenous variable in the extent-of-hedging equation, and (2) the extent of hedging is an endogenous variable in the extent-of-smoothing with AACs equation.

We evaluate firm i 's year t decision whether to hedge as follows:

$$Hedgers_{it} = \zeta_0 + \zeta_1 Explrisk_{it} + \zeta_2 FullCost_i + \zeta_3 HdgControls_{it} + \xi_{it} \quad (1)$$

and then, given that hedging occurs, we simultaneously assess the decisions about the extent of hedging and the extent of smoothing with AACs, using the following two equations:

$$Hedging\ ratio_{it} = \gamma_0 + \gamma_1 PredAAC\ smoothing\ ratio_{it} + \gamma_2 Explrisk_{it} + \gamma_3 FullCost_i + \gamma_4 HdgControls_{it} + \gamma_5 InvMills_{it} + \omega_{it} \quad (2)$$

$$AAC\ Smoothing\ ratio_{it} = \phi_0 + \phi_1 PredHedging\ ratio_{it} + \phi_2 Explrisk_{it} + \phi_3 FullCost_i + \phi_4 AACControls_{it} + \phi_5 InvMills_{it} + \kappa_{it} \quad (3)$$

where:

$Hedgers_{it}$ = 1 if firm i holds a nonzero derivative position at fiscal t year-end, and 0 otherwise;

$Hedging\ ratio_{it}$ = quantity of oil and gas production that firm i hedged at fiscal t year-end, scaled by quantity of year t production;

$PredHedging\ ratio_{it}$ = predicted value of $Hedging\ ratio_{it}$ from the first stage of two-stage least squares (2SLS);

$AAC\ smoothing\ ratio_{it}$ = smoothing with abnormal accruals ratio = standard deviation of firm i 's quarterly income before abnormal accruals and extraordinary items in year t divided by standard deviation of firm i 's quarterly income before extraordinary items in year t ;

$PredAAC\ smoothing\ ratio_{it}$ = predicted value of $AAC\ smoothing\ ratio_{it}$ from the first stage of 2SLS;

$Explrisk_{it}$ = exploration risk = firm i 's year t score from a factor analysis of two exploration risk proxies, exploration expenditures, and Sunder's (1976) variance;

$FullCost_i$ = 1 if firm i uses full cost, and 0 if it uses successful efforts;

$HdgControls_{it}$ and $AACControls_{it}$ = additional control variables for the hedging and smoothing with abnormal accruals equations, respectively; and

$InvMills_{it}$ = inverse Mills ratio = self-selection adjustment from estimating Equation (1).

We employ an estimation approach based on Cragg's (1971) self-selection model (see also Heckman 1979). We model the initial decision of whether to hedge as a binomial probit regression, estimated using all firm-years in the sample. Next, we consider only firm-years with hedging, and we simultaneously estimate regressions for the extent of hedging and extent of smoothing with AACs. Because a firm's decision whether to hedge is not random, selectivity bias can cause ξ_{it} , ω_{it} , and κ_{it} in Equations (1)–(3) to be correlated, which can lead to biased estimates of the regression parameters. Thus, we incorporate the adjustment for self-selection (the inverse Mills ratio) from the estimated probit model into the hedging Equation (2) and the AAC smoothing Equation (3). We then estimate Equations (2)–(3) using 2SLS, which is valid asymptotically and subject to other limitations (Kennedy 1998, chapter 10; Holthausen et al. 1995, 296–297). In the first stage, we regress each endogenous variable (*AAC smoothing ratio* and *Hedging ratio*) on the exogenous variables and then compute predicted values of the hedging and smoothing ratios. We label the resulting predicted variables *PredHedging ratio* and *PredAAC smoothing ratio* and use them as endogenous variables in the second stage of the estimation along with the exogenous variables.

We test our basic hypothesis by estimating Equations (2) and (3). If the extent of hedging and the extent of smoothing with AACs are substitutes, and if managers make both decisions simultaneously, then the coefficient on *PredAAC smoothing ratio*_{it} in Equation (2) should be negative ($\gamma_1 < 0$) and the coefficient on *PredHedging ratio*_{it} in Equation (3) should be negative ($\phi_1 < 0$). Note that if managers make the decisions sequentially, then the negative relation will occur in Equation (2) or Equation (3), but not in both, and if managers do not use hedging and smoothing with AACs as substitutes, then neither γ_1 nor ϕ_1 will be negative.

Sample Selection

We begin the empirical analysis in 1993 because few firms made voluntary disclosures about derivatives prior to that time. Because we can identify their inherent risks, we focus on oil and gas exploration and producing firms (SIC code 1311) and exclude large, vertically integrated firms that explore, extract, transport, refine, and distribute oil and gas products. We initially identified 163 companies in the 1996 Compustat annual files. We deleted (1) thirteen firms that were undergoing bankruptcy or liquidation proceedings, or experiencing going-concern problems, (2) three firms that were subsidiaries of other firms in the sample, (3) four firms that switched from full cost to successful efforts accounting, or vice versa, during the period 1993 to 1996, and (4) four firms for which financial statements were not available for the period. The remaining sample is 139 firms from the period 1993–1996. However, 124 firm-years lack data to compute the hedging ratio, an additional 182 firm-years lack data to calculate the explanatory variables for the hedging regressions, and another 14 firm-years lack data to compute the variables for the smoothing with AACs regression. The final sample is 236 firm-years. In untabulated results, we compare the final sample with the deleted firm-years that have at least some of the required data, and find the following significant differences: the excluded firm-years reflect smaller firms, a lower

occurrence of hedging, less extensive hedging when hedging occurs, and less extensive smoothing with AACs than the firm-years included in our sample.

Dependent Variables

Table 1 provides definitions and data sources for all of the study's variables. We use annual report or 10-K disclosures of year-end commodity derivative positions to document the occurrence of hedging (*Hedgers_{it}*) and the extent of hedging (*Hedging ratio_{it}*). The numerator of *Hedging ratio* is the quantity of production hedged, and the denominator is the quantity of production, which reflects the firm's exposure to oil price risk (Haushalter 2000).^{2,3} Appendix A offers an example of derivative disclosures and illustrates the computation of *Hedging ratio*.

Our measure of smoothing with abnormal accruals, *AAC smoothing ratio_{it}*, equals the standard deviation of firm *i*'s year *t* quarterly earnings before abnormal accruals divided by the standard deviation of its year *t* quarterly earnings; i.e., σ_{EBAAC}/σ_E (Hunt et al. 1997). Values of *AAC smoothing ratio* in excess of 1 indicate more variability in earnings before abnormal accruals than in earnings after abnormal accruals, consistent with smoothing via AACs.⁴ We compute *AAC smoothing ratio* for each firm-year based on quarterly data. Specifically, we define (1) earnings as income before extraordinary items and (2) earnings before abnormal accruals as operating cash flows plus normal accruals; we scale each by total assets of the previous quarter. We measure quarterly operating cash flows following Han and Wang (1998, notes 10 and 11), and compute quarterly normal accruals by adapting the modified cross-sectional Jones model (Dechow et al. 1995) to include interactions of each explanatory variable with a dummy variable that equals 1 (0) if a firm uses (does not use) full cost to account for exploration costs. We discuss the adapted accruals model and the motivation for it below.⁵

² Barton (2001) uses lagged total assets as the denominator in his derivatives variable. However, Tufano (1996) and Wong (2000) argue that a firm's hedge position should be evaluated with respect to what is being hedged. In addition, Barton (2001, note 4) uses the notional amount of derivatives reported under SFAS No. 119 (FASB 1994) as the numerator in his derivatives variable. Wong (2000, 393) notes that SFAS No. 119 aggregated notional-amount disclosures do not distinguish clearly whether a firm has assumed a long or short position. We do not face that problem because oil and gas producing firms are naturally long in oil price risk. However, because we rely on commodity derivative data that firms voluntarily disclose, our analysis excludes firms that hedge oil price risk but do not disclose that fact. Also, we base our measure on a firm's net derivative position. Although the net position reflects the derivatives firms held for trading purposes, this measurement error should not materially affect our analysis because only 7 percent of our firm-years use derivatives for trading purposes.

³ The numerator of *Hedging ratio* is the total notional quantity of oil and gas hedged, which we determine by aggregating across various derivative types. However, options are one-sided contracts, whereas swaps, forwards, and futures are two-sided contracts (there are receivable/payable implications regardless of whether the oil price rises above or falls below the strike price of the derivative). We find that managers use options in only 10 percent of our sample firm-years. Hence, aggregating across instrument types is unlikely to induce significant measurement error in our hedging ratio variable.

⁴ Barton (2001) uses the absolute value of abnormal accruals. This is an indirect measure that is subject to the limitation that larger absolute values of abnormal accruals may not always result in smoother earnings.

⁵ We examined in depth one-third of our sample firms' financial statements and notes to identify the common types of accounts reflecting accruals. We find that operating current assets and liabilities include accounts receivable, inventories, prepaid expenses, payables, accrued expenses, and unearned revenues. Since these accounts are typical of companies in general, the modified Jones model would seem to be appropriate in our oil and gas setting. Operating accounts that reflect oil and gas exploration and production activities include (1) receivables (payables) from deliveries of natural gas due to excess production (underproduction) where revenue is based on a firm's working interest or entitlement in a field's production, and (2) unearned oil and gas revenues where firms receive advance payments for deliveries of future production at a certain price. In Section IV, we assess the sensitivity of the results to measurement error in estimating abnormal accruals (Bernard and Skinner 1996), including the effect of special items and of acquisitions and divestitures (Collins and Hribar 2000).

TABLE 1
Variable Definitions and Data Sources

Dependent/Endogenous Variables:

$Hedgers_{it}$ = indicator variable that equals 1 if firm i holds a nonzero derivative position at fiscal t year-end, and 0 otherwise (Source: 10-K, Annual Report).

$Hedging\ ratio_{it}$ = quantity of oil and gas production firm i hedged at fiscal t year-end, scaled by quantity of year t production (Source: 10-K, Annual Report; see Appendix A).

$AAC\ smoothing\ ratio_{it}$ = smoothing with abnormal accruals ratio = standard deviation of firm i 's quarterly income before abnormal accruals and extraordinary items in year t divided by standard deviation of its quarterly earnings before extraordinary items in year t , i.e., σ_{EBAAC}/σ_E (Hunt et al. 1997). The following summarizes key variables used or computed in quarter q for firm i (variables are scaled by total assets of the previous quarter, Compustat quarterly data item #44):

$$\begin{aligned} EBAAC_{iq} &= CFO_{iq} + NAC_{iq} \\ TotalAC_{iq} &= EBEI_{iq} - CFO_{iq} \\ CFO_{iq} &= WCO_{iq} - \Delta AR_{iq} - \Delta INV_{iq} - \Delta OCA_{iq} + \Delta AP_{iq} \\ &\quad + \Delta TP_{iq} + \Delta OCL_{iq} \end{aligned}$$

where $EBAAC_{iq}$ = earnings before abnormal accruals; CFO_{iq} = cash flows from operations; NAC_{iq} = normal accruals computed using an adaptation of the modified cross-sectional Jones model that incorporates a firm's choice of full cost or successful efforts accounting (Equation [4] in the text); $TotalAC_{iq}$ = total accruals; $EBEI_{iq}$ = income before extraordinary items (Compustat quarterly data item #69); WCO_{iq} = working capital from operations (following Han and Wang 1998, notes 10 and 11) = $EBEI_{iq} \pm$ noncash, nonoperating expenses and revenues (Compustat quarterly data item #76 + #77 + #78 + #79 + #80 + #81 + #102); ΔAR_{iq} = change in accounts receivable (Compustat quarterly data item #37); ΔINV_{iq} = change in inventory (#38); ΔOCA_{iq} = change in other current assets (#39); ΔAP_{iq} = change in accounts payable (#46); ΔTP_{iq} = change in taxes payable (#47); ΔOCL_{iq} = change in other current liabilities (#48). We adjust data items reported on a cumulative basis in Compustat to reflect quarterly values, and changes in working capital accounts are the differences between the current amount and the prior quarter's amount.

Explanatory Variables—Hedging and Smoothing with AACs Regressions:

$Explrisk_{it}$ = exploration risk = score from a factor analysis of: (1) firm i 's year t oil and gas exploration expenditures, and (2) the firm's year t variance of net operating cash flows, assuming it is in steady-state (Sunder 1976), both scaled by the firm's year-end reserve values. We identify from factor analysis one factor having an eigenvalue greater than 1 (Source: 10-K, Annual Report; Appendix B details the computation of Sunder's variance).

$FullCost_i$ = indicator variable that equals 1 if firm i uses the full cost method, and 0 if it uses successful efforts (Source: Compustat annual footnote 31, 10-K, Annual Report).

$Leverage_{it}$ = long-term debt scaled by market value of equity, both measured at fiscal t year-end (Compustat annual data item #9/[#199 × #25]).

M/B_{it} = market-to-book ratio = ratio of market value of equity to book value of equity of firm i , each measured at fiscal t year-end (Compustat annual data item [#199 × #25]/[#60 + #30]).

(Continued on next page)

TABLE 1 (Continued)

Tax_{it} = indicator variable that equals 1 if firm i is profitable (i.e., income before extraordinary items > 0) in year t and has NOL tax carryforwards at fiscal t year-end, and 0 otherwise (Compustat annual data item #18 and #52, respectively).

$MgrlOwn_{it}$ = managerial ownership = percentage of firm i 's shares held by insiders in year t (Source: Compact D-SEC).

$Stock\ options_{it}$ = number of exercisable stock options managers and employees hold, scaled by number of shares outstanding, both measured at fiscal t year-end (Source: 10-K, Annual Report).

$InstitOwn_{it}$ = percentage of firm i 's total shares outstanding held by institutions in year t (Source: Compact D-SEC).

$Firm\ size_{it}$ = log of market value of equity of firm i at fiscal t year-end (Compustat annual data item #199 $\times$ #25).

Explanatory Variables Unique to Hedging Regressions:

σRET_{it} = standard deviation of returns = computed using firm i 's monthly returns over fiscal year t (Source: CRSP).

$Production\ exposed_{it}$ = proportion of firm i 's year t production exposed to basis risk because it is produced in locations other than Arkansas, Kansas, Louisiana, Oklahoma, and Texas (Source: 10-K, Annual Report).

$Intl\ production_{it}$ = international production = indicator variable that equals 1 if firm i has production in year t at international locations, and 0 otherwise (Source: 10-K, Annual Report).

$O\&G\ Production_{it}$ = oil and gas production = percentage of firm i 's fiscal year t sales from oil and gas production (Source: 10-K).

$Cash_{it}$ = firm i 's cash scaled by its market value of equity at fiscal t year-end (Compustat annual data item #1/[#199 $\times$ #25]).

Explanatory Variables Unique to Smoothing with AACs Regression:

$DivPayout_{it}$ = dividend payout ratio = dividends per share to common shareholders of firm i in fiscal year t divided by earnings per share before extraordinary items in year t (Compustat annual data item #26/#58).

$MarkToMarket_{it}$ = indicator variable that equals 1 if firm i trades in derivatives in year t , and 0 otherwise (Source: 10-K, Annual Report).

Controls: Exploration Risk and Accounting for Exploration Costs

Our proxy for exploration risk ($Explrisk_{it}$) is the score from a factor analysis of (1) a firm's annual oil and gas exploration expenditures (Malmquist 1990), and (2) the firm's variance of net operating cash flows, assuming it is in steady-state (Sunder 1976). In Appendix B we summarize the theoretical derivation of Sunder's variance, note the assumptions we make to estimate it, and illustrate its computation. We scale exploration costs and Sunder's variance by the firm's year-end reserve values and identify from factor analysis one factor having an eigenvalue greater than 1 (not shown). This factor retains 85 percent of the variation in the input variables. Using the estimated weights from the factor analysis, we linearly combine the two input variables to derive factor scores for each firm-year. Higher factor scores indicate more exposure to exploration risk and hence greater concern about earnings volatility.

The full cost method of accounting for exploration costs views an entire drilling area as an asset, and firms capitalize and amortize all exploration costs against future earnings.

Under successful efforts, however, only productive wells are assets. A firm using successful efforts expenses the costs of a dry well in the period it determines the well is uneconomic, rather than amortizing all exploration costs over a longer period of time. Hence, full cost typically generates a smoother time-series of earnings than successful efforts,⁶ so full cost firms may be less inclined to hedge or smooth with AACs. However, full cost firms may be fundamentally different from successful efforts firms (Malmquist 1990), their earnings streams may reflect such fundamental differences (Sunder 1976), and they may face different levels of overall risk. It is thus unclear whether (or how) the use of full cost or successful efforts is associated with current period hedging and smoothing with AACs. We control for firms' use of full cost or successful efforts by including the indicator variable *FullCost_i* in both the hedging and the smoothing-with-AACs equations, without making a directional prediction.⁷ Furthermore, given the possibility of substantive differences between full cost and successful efforts firms and the two methods' differential effects on the time-series of earnings and also on property, plant, and equipment, and therefore total assets, we adapt the modified Jones model in estimating accrual components to account for a firm's choice of full cost or successful efforts. Specifically, we estimate the following model:

$$\begin{aligned} TotalAC_{iq}/TA_{iq-1} = & a_1(1/TA_{iq-1}) + a_2[(\Delta Rev_{iq} - \Delta Rec_{iq})/TA_{iq-1}] \\ & + a_3(PPE_{iq}/TA_{iq-1}) + a_{1a}(1/TA_{iq-1})D_i \\ & + a_{2a}[(\Delta Rev_{iq} - \Delta Rec_{iq})/TA_{iq-1}]D_i \\ & + a_{3a}(PPE_{iq}/TA_{iq-1})D_i + v_{iq} \end{aligned} \quad (4)$$

where:

TotalAC_{iq} = total accruals for firm *i* in quarter *q*, measured as income before extraordinary items – cash flows from operations;

TA_{iq-1} = firm *i*'s total assets in quarter *q* – 1;

ΔRev_{iq} = firm *i*'s change in revenues from quarter *q* – 1 to *q*;

ΔRec_{iq} = firm *i*'s change in receivables from quarter *q* – 1 to *q*;

PPE_{iq} = firm *i*'s gross property, plant, and equipment in quarter *q*;

D_i = 1 (0) if firm *i* uses the full cost (successful efforts) method; and

v_{iq} = residual from ordinary least squares (OLS) estimation = abnormal accruals.

Other Control Variables

Equations (1)–(3) control for a number of additional factors we expect to affect hedging and/or smoothing with AACs.

⁶ However, full cost induces greater earnings variability when sharp oil-price declines necessitate reserve write-downs. Sharp annual price declines did not occur in our sample period (Rajgopal 1999).

⁷ Treating the full cost/successful efforts (FC/SE) choice as exogenous is a limitation of our analysis. We perform sensitivity checks by interacting *FullCost* with variables in the hedging and smoothing-with-AACs models. From the set of firms with the necessary data, only four firms (3 percent) changed their FC/SE choice over our test period, and only six firms switched during the preceding 12 years. Firms likely find it less costly to smooth earnings by using AACs or hedging than by changing their accounting methods for exploration costs. Furthermore, it seems unlikely that the original determinants of firms' FC/SE choices remained constant since firms made their choices, and the results in Table 2, Panel C indicate that full cost and successful efforts users differ in ways that Malmquist (1990) did not identify. It is therefore beyond the scope of our study to model the determinants of the FC/SE choice for our oil and gas firms.

Additional Factors Expected to Affect Both Hedging and Smoothing with AACs

Financial leverage. The greater a firm's debt, the more likely it will hedge. Debt contracts typically constrain firms to reduce the probability of financial distress, and hedging mitigates extremely negative cash flows (Geczy et al. 1997; Graham and Rogers 1999). Smoothing with AACs also reduces the likelihood of reporting severe losses and thus of technical default. We use long-term debt scaled by market value of equity as our *Leverage* proxy.

Investment opportunity set. The more growth opportunities available, the more likely a firm will hedge cash flows to assure the availability of funds. Growth opportunities also provide an incentive to smooth earnings using AACs because earnings volatility reflects firm risk (Beaver et al. 1970) and thus potentially adversely affects the cost of the capital needed to fund investment projects. We use the market-to-book (*M/B*) ratio to proxy for growth opportunities.

Additionally, Froot et al. (1993) argue that hedging mitigates underinvestment by reducing a firm's dependence on, and costs of, external financing. If external financing is more costly than internally generated funds, then a firm that does not hedge to reduce the volatility in its operating cash flows may underinvest if it is too costly to raise funds externally. Hedging thus allows the firm to avoid unnecessary fluctuations in either investment spending or externally obtained financing. We proxy for the costs of underinvestment using the interaction of growth opportunities and debt financing, $M/B \times Leverage$ (Geczy et al. 1997), and predict a positive relation between this interaction and both hedging and smoothing with AACs.

Income taxes. Graham and Smith (1999) show that firms with existing net operating loss carryforwards (NOLs) have an incentive to hedge if they expect to be profitable. The incentive derives from the asymmetric tax treatment of profits and losses and limitations to sell or immediately use tax preference items, such as NOLs (Smith and Stulz 1985; Graham and Rogers 1999). The indicator variable *Tax* equals 1 in a year when the firm is profitable and has NOL tax carryforwards, and so has a tax incentive to smooth, either by hedging or by using AACs.

Managerial wealth and risk. If risk-averse managers cannot diversify firm-specific risks or if they believe that the market perceives lower earnings volatility as reflecting lower firm riskiness, then they have more incentive to hedge or smooth with AACs, the larger their holdings in their firms' stocks (Smith and Stulz 1985; Guay 1999). On the other hand, owner-controlled firms provide additional monitoring and tend to manage earnings less (Warfield et al. 1995). We proxy managerial ownership (*MgrlOwn*) by using the percentage of firm *i*'s shares held by insiders, but we do not predict the direction of its association with hedging and smoothing with AACs.

Managers holding more stock options have less incentive to dampen volatility if the value of options increases in volatility (Tufano 1996). This incentive is more salient for exercisable options (Schrand and Unal 1998; Haushalter 2000). Thus, we expect hedging and smoothing with AACs to be inversely related to *Stock options*, the number of exercisable stock options managers and employees hold, scaled by the number of shares outstanding, both as of year-end.⁸

⁸ Sensitivity of options to returns or returns volatility is arguably a better proxy for stock option based earnings management incentives. However, computing such sensitivity requires extensive information, such as time to maturity and exercise price of new and previously granted options. Footnote data do not provide full disclosure of previously granted options, and *Execucomp* includes complete data for only 25 sample firms.

Institutional ownership. External parties cannot observe managerial quality, making it difficult to disentangle profits due to managerial ability from profits due to exogenous shocks (DeMarzo and Duffie 1995). The less the external monitoring of the firm, the greater managers' incentives to hedge cash flow volatility, and to smooth earnings with AACs, to facilitate the market's assessment of their skills. We use the extent of institutional ownership (*InstitOwn*) to proxy for the degree of external monitoring, based on the assumption that more extensive institutional ownership leads to monitoring that in turn reduces information asymmetry between investors and managers (Geczy et al. 1997). These arguments suggest a negative relation between *InstitOwn* and both *Hedging ratio* and *AAC smoothing ratio*. On the other hand, external monitoring likely increases pressure on managers to dampen volatility—i.e., to make earnings more predictable (Levitt 1998; Loomis 1999), suggesting a positive relation between extent of institutional ownership and both hedging and smoothing with AACs. Accordingly, we do not sign the predicted association.⁹

Firm size. Larger firms enjoy the economies of scale to obtain expertise and lower average transaction costs needed to hedge effectively (Mian 1996; Geczy et al. 1997; Haushalter 2000). We use *Firm size*, defined as the log of a firm's market value of equity at year-end, as our proxy for scale, and predict a positive relation between firm size and hedging.

With regard to smoothing with AACs, note that larger firms are also subject to more external monitoring, which constrains managers' ability to smooth earnings with AACs. However, larger firms are followed by more analysts (Bhushan 1989) and arguably face more pressure to report more predictable earnings (Fox 1997). Thus, we do not predict the direction of the relation between *Firm size* and *AAC smoothing ratio*, in contrast to the predicted positive link between *Firm size* and hedging.

Year indicators. We include a dummy variable for each year (except 1996) to proxy for changes in unspecified macroeconomic factors, which are cross-sectional constants (such as oil prices). We thus estimate fixed-effects models.

Factors to Discriminate between Hedging and Smoothing with Abnormal Accruals

Hedging: Cost of capital. We use the standard deviation of the firm's monthly returns over the fiscal year (σRET) as a proxy for the cost of capital associated with cash flow volatility (Minton and Schrand 1999). Firms have a greater incentive to hedge, the greater the σRET . However, we can observe σRET only *after* the firm has hedged. To the extent σRET incorporates the effects of hedging, and hedging has successfully reduced σRET to below σRET of non-hedgers, we would expect lower σRET to result from more extensive hedging, and therefore predict a negative sign. Also in our regression model, σRET captures stock return volatility incremental to that contributed by exploration risk because exploration risk is a separate independent variable in the hedge/no hedge and *Hedging ratio* regressions.

Hedging: Basis risk. Haushalter (2000) reports that in Arkansas, Kansas, Louisiana, Oklahoma, and Texas, spot prices for oil and gas are highly correlated with the two benchmark grades of oil and gas (West Texas Intermediate and Henry Hub) on which most derivative contracts are written. Production in these locations faces relatively low basis risk, making hedging effective. On the other hand, hedging will be less effective for production in other locations, and thus managers will be less likely to hedge. Our proxy for basis risk,

⁹ Including indicator variables for exchange listing and for Big 6 auditor as additional proxies for external monitoring does not affect the study's inferences.

Production exposed, is the proportion of the firm's annual production *not* located in Arkansas, Kansas, Louisiana, Oklahoma, and Texas, and we predict a negative relation between *Production exposed* and hedging. We also control for *Intl production*, which equals 1 if a firm has oil and gas production at international locations, and 0 otherwise, because it is more difficult to identify derivatives with value changes that are highly correlated with changes in the value of foreign production.

Hedging: Substitutes. Diversification of operations is a possible hedging substitute because shocks in one line of business may offset shocks in other lines of business. We expect managers are more likely to hedge, the less diversified the firm's operations (i.e., the larger the portion of a firm's revenues derived from oil and gas production, as proxied by the percentage of annual sales from oil and gas production, *O&G Production*).

If managers hedge to dampen cash flow volatility, then the availability of cash should reduce the need, if not substitute, for hedging (Haushalter 2000). We expect that the more cash on hand, the less managers will hedge. *Cash* is defined as cash scaled by year-end market value of equity.

Smoothing with AACs: Dividend payout ratio. Volatility in earnings affects firms' ability to pay dividends because dividend restrictions in bond covenants are usually based on earnings realizations (Smith and Warner 1979). We compute dividend payout ratio (*DivPayout*) as dividends to common shareholders divided by earnings before extraordinary items, and predict a positive relation with *AAC smoothing ratio*.

Smoothing with AACs: Accounting for derivatives used for trading purposes. Firms that use derivatives for trading purposes use mark-to-market accounting, which can induce earnings volatility. If firms engaged in derivative trading also are concerned about earnings volatility, then we expect them to use AACs to smooth earnings. We use a dummy variable, *MarkToMarket*, which equals 1 if a firm trades in derivatives in year *t*, and 0 otherwise.

IV. DATA AND RESULTS

Sample Description

Panel A of Table 2 reports descriptive statistics for the full sample. With regard to the dependent variables, 44 percent of firm-years hedge oil price risk with derivatives (*Hedgers*), and firms on average hedge 33 percent of production (*Hedging ratio*). There is considerable variation in the proportion of production hedged (the coefficient of variation equals 3.58), and more often than not oil and gas producers do not hedge, consistent with Haushalter (2000). In the subset of firm-years with hedging (not shown), on which we conduct the study's main analyses, the mean (median) proportion of production hedged is 66 percent (30 percent).

Across all firm-years, the mean *AAC smoothing ratio* is 4.20 (median = 2.17). In untabulated results, a t-test rejects the null hypothesis that the mean equals 1 ($t = 9.71$) and *AAC smoothing ratio* values exceed 1 in 64 percent of firm-years. In firm-years with hedging, the mean *AAC smoothing ratio* equals 5.41 (median = 3.05); the mean reliably exceeds 1, as do 72 percent of individual *AAC smoothing ratio* values (not shown). These results are consistent with pervasive and nontrivial smoothing.

Turning to the independent variables, firms in our sample are almost evenly split between those using full cost and those using successful efforts.¹⁰ There is substantial variation across firm-years in *Explrisk* (coefficient of variation = 4.44) and in growth opportunities,

¹⁰ Prior studies find that large firms typically use successful efforts (Foster 1980), but there is no significant relation between *Firm size* and *FullCost* in our sample. The likely explanation for the difference is that our sample includes oil and gas exploration and drilling firms, and excludes large, vertically integrated companies.

TABLE 2
Descriptive Statistics and Univariate Analyses of the Sample of Oil and Gas Producing Firms
(SIC Code 1311) over the Period 1993–1996

Panel A: Descriptive Statistics of the Full Sample (n = 236 firm-years)

<i>Variable</i>	<i>Mean</i>	<i>Median</i>	<i>Standard Deviation</i>	<i>1st Quartile</i>	<i>3rd Quartile</i>
Dependent variables:					
<i>Hedgers</i>	0.44	0	0.50	0	1
<i>Hedging ratio</i>	0.33	0	1.17	0	0.26
<i>AAC smoothing ratio</i>	4.20	2.17	5.09	0.58	6.25
Common explanatory variables:					
<i>Explrisk</i>	0.13	0.01	0.58	0.00	0.05
<i>FullCost</i>	0.50	0	0.50	0	1
<i>Leverage</i>	0.42	0.24	0.65	0.07	0.47
<i>M/B</i>	1.45	1.34	5.05	0.83	2.02
<i>Tax</i>	0.15	0	0.36	0	1
<i>MgrlOwn</i> (percent)	19.13	7.51	25.45	1.4	25.76
<i>Stock options</i>	0.07	0.04	0.12	0.02	0.08
<i>InstitOwn</i> (percent)	33.19	28.98	28.46	1.50	59.32
<i>Firm size</i> (millions of \$)	522.64	65.03	1125.30	23.70	333.32
Variables for hedging equations:					
<i>σRET</i>	0.12	0.10	0.07	0.08	0.14
<i>Production exposed</i> (percent)	24.22	6	33.91	0	33
<i>Intl production</i>	0.36	0	0.48	0	1
<i>O&G Production</i> (percent)	89.11	100	17.64	84	100
<i>Cash</i>	0.11	0.05	0.18	0.01	0.12
Variables for AAC smoothing equation:					
<i>DivPayout</i>	0.72	0	1.68	0	0.40
<i>MarkToMarket</i>	0.07	0	0.26	0	0

(Continued on next page)

TABLE 2 (Continued)

Panel B: Differences between Hedgers ($n = 103$ firm-years) and Non-Hedgers ($n = 133$ firm-years)

Variable	Hypothesized Sign ^a	Medians		Wilcoxon Rank Sum Test Higher Score	p-value ^b	t-test of Differences in Means	
		Hedger (H)	Non-Hedger (NH)			t-statistic	p-value
Endogenous variable: AAC smoothing ratio	H > < NH	3.05	1.68	Hedger	0.00	1.86	0.00
Common explanatory variables:							
ExplRisk	H > NH	0.01	0.00	Hedger	0.00	1.57	0.06
FullCost	H > < NH	0	0	Non-hedger	0.91	-0.11	0.91
Leverage	H > NH	0.32	0.14	Hedger	0.00	2.89	0.00
M/B	H > NH	1.30	1.31	Non-hedger	0.45	1.42	0.92
Tax	H > NH	0	0	Hedger	0.15	1.01	0.15
MgrtOwn (percent)	H > < NH	8.23	7.51	Non-hedger	0.57	0.76	0.44
Stock options	H < NH	0.04	0.03	Hedger	0.35	-1.12	0.87
InstitOwn (percent)	H > < NH	51.67	3.48	Hedger	0.00	8.04	0.00
Firm size	H > NH	5.16	3.31	Hedger	0.00	4.21	0.00
Variables for hedging equations:							
σ_{RET}	H < NH	0.09	0.11	Non-hedger	0.00	-3.17	0.00
Production exposed (percent)	H < NH	6	2	Hedger	0.23	-0.32	0.63
Int'l production	H < NH	0	0	Hedger	0.47	0.37	0.47
O&G Production (percent)	H > NH	98	100	Non-hedger	0.11	0.04	0.52
Cash	H < NH	0.03	0.06	Non-hedger	0.00	-3.17	0.00
Variables for AAC smoothing equation:							
DivPayout	H > < NH	0	0	Hedger	0.03	0.51	0.60
MarkToMarket	H > < NH	0	0	Hedger	0.00	2.44	0.00

(Continued on next page)

TABLE 2 (Continued)

Panel C: For the Sample of Firm-Years with Hedging, Differences between Full Cost ($n = 55$ firm years) and Successful Efforts ($n = 48$ firm years)^c

Variable	Medians		Wilcoxon Rank Sum Test		t-test of Difference in Means	
	FC Firm	SE Firm	Higher Score	p-value ^d	t-statistic	p-value
Endogenous variables:						
AAC smoothing ratio	6.13	2.68	FC firm	0.00	2.66	0.00
Hedging ratio	0.27	0.26	FC firm	0.79	-0.09	0.92
Common explanatory variables:						
Explrisk	0.01	0.00	FC firm	0.23	1.21	0.22
Leverage	0.28	0.37	SE firm	0.04	-1.86	0.06
M/B	1.35	1.26	SE firm	0.75	-0.94	0.38
Tax	0	0	SE firm	0.71	-0.37	0.70
MgrlOwn	14.47	7.85	FC firm	0.32	0.47	0.63
Stock option	0.05	0.04	FC firm	0.05	2.34	0.02
InstitOwn	58.46	51.67	FC firm	0.89	0.24	0.81
Firm size	5.08	5.38	SE firm	0.22	-1.47	0.14
Variables for hedging equations:						
σ RET	0.10	0.09	FC firm	0.05	1.24	0.21
Production exposed	10.50	4.81	FC firm	0.07	2.24	0.01
Intl production	0	0	SE firm	0.53	-0.61	0.54
O&G Production	100	91.85	FC firm	0.03	2.03	0.04
Cash	0.04	0.03	FC firm	0.92	-1.05	0.29
Variables for AAC smoothing equation:						
DivPayout	0	0	SE firm	0.00	0.24	0.81
MarkToMarket	0	0	SE firm	0.25	-1.12	0.26

^a Hypothesized sign is based on the discussion in Section III for Equations (1)–(3). Where there is no prediction for a variable regarding the hedge/no hedge decision, we include the variable for completeness.

^b p-values are one-tailed if there is a directional prediction; two-tailed otherwise.

^c No predictions were developed for full cost and successful efforts methods in the full sample.

^d p-values are two-tailed.

Coefficients are in **bold** print if in the hypothesized direction when both the t-test and Wilcoxon tests are significant at $p < 0.10$.

M/B (coefficient of variation = 3.48), and firms are leveraged to a considerable degree (mean *Leverage* = 0.42, median = 0.24). Fifteen percent of firm-years have an NOL carryforward and are profitable. Institutional investors own approximately 30 percent of the shares of our sample firms, whereas managers hold, on average, 19 percent (median = 7.5 percent) of their firms' shares and have exercisable options for an additional 7 percent (median = 4 percent). The mean market value of equity in the sample is \$522 million, which is substantially greater than the median of \$65 million. This skewness prompts us to use the log of market value of equity as our *Firm size* variable in the subsequent empirical analysis.¹¹

Of the variables that are likely determinants of hedging alone, average variability of returns (σRET) is approximately 12 percent per firm-year. The mean level of *Production exposed* to basis risk is 24 percent, while 36 percent of firm-years reflect production at international locations. There is little diversification of operations, since the mean *O&G Production* is 89 percent and the median is 100 percent. *Cash* availability averages 0.11 (median = 0.05) of the market value of equity.

As for explanatory variables identified as determinants only of smoothing with AACs, in most firm-years managers pay no dividends, but there are a few cases of very large *DivPayout* values.¹² Managers use derivatives for trading purposes in only 7 percent of firm-years.¹³

Univariate Results

For descriptive purposes, we present univariate comparisons of two partitions of the sample in Panels B and C of Table 2, using one-tailed tests of differences where we have a directional prediction, and two-tailed tests otherwise. In Panel B, we compare the 103 sample firm-years with hedging to the 133 firm-years without hedging. Larger and highly leveraged firms are more likely to hedge, as are firms for which institutions hold a relatively high proportion of the firm's shares. Hedgers on average have lower levels of both *Cash* and σRET . Hedging is also associated with higher average levels of both exploration risk and smoothing with AACs. Untabulated results reveal that the Spearman correlation between *Hedging ratio* and *AAC smoothing ratio* for the full sample is significantly positive ($\rho_s = 0.138$; $p = 0.02$), suggesting that firms use both hedging and smoothing with AACs to manage volatility. However, within the sample of 103 firm-years with hedging, we find negative correlations between *Hedging ratio* and *AAC smoothing ratio* (Pearson = -0.18 , $p = 0.07$; Spearman = -0.12 , $p = 0.10$), consistent with managers who do decide to hedge using these smoothing mechanisms as substitutes. These univariate results suggest that the decision to hedge is associated with a greater level of smoothing with abnormal accruals, but that once the firm decides to hedge, the amount hedged is inversely related to the amount of smoothing with abnormal accruals, consistent with a trade-off at the margin. Our simultaneous equation design allows for substitutions between these two smoothing tools at the margin and controls for factors that affect cross-sectional differences in incentives for smoothing.

Our other univariate comparison divides the 103 hedging firm-years into full cost and successful efforts usage (see Table 2, Panel C). Full cost is associated with lower leverage, greater managerial holdings of stock options, greater exposure to basis risk, less diversified

¹¹ Other variables also reflect skewness. We re-run all regressions after down-weighting influential observations (Belsley et al. 1980) using the *RWEIGHT* function in SAS. Our primary inferences are unaffected.

¹² Our inferences are unchanged when we use $\log(1 + DivPayout)$.

¹³ About 40 percent of sample firms' shares trade on the NYSE, and the Big 6 audit 81 percent of the firm-years.

operations, and higher levels of smoothing with AACs. Similar to Geczy et al. (1999), we find no relation between the extent of hedging and the choice of full cost or successful efforts.

Few correlations between variables (in the full sample and in the subset of firm-years with hedging) exceed 0.30 (not shown). Not surprisingly, two exceptions are that the correlation between M/B and $M/B \times Leverage$ is 0.94, and the correlation between the log of firm size and *InstitOwn* is 0.60. In our empirical analysis we estimate alternative models that, in turn, include and exclude highly correlated variables to address concerns about potential problems due to multicollinearity.

Results of Primary Hypothesis Tests

Our primary results are based on the analysis presented in Table 3. We initially estimate a binomial probit model to identify the determinants of the decision of whether to hedge, and to extract the inverse Mills ratio. We then use 2SLS in the sample of firm-years in which hedging occurs. We determine predicted values of the endogenous variables and include them along with the exogenous variables and the inverse Mills ratio in second-stage regressions. We report one-tailed p-values unless the prediction is nondirectional.

The results of the hedge/no hedge (1/0) probit appear under the heading of *Hedgers*. The positive coefficient on *Explrisk* ($p = 0.10$) indicates that firms with more exploration risk are more likely to hedge. $M/B \times Leverage$ has a significantly positive coefficient, consistent with a higher probability of hedging the greater the costs of underinvestment, as reflected in the interaction between growth opportunities and debt financing. Contrary to expectations, the coefficient on M/B is negative. The high correlation between the two growth-opportunities variables likely explains the unexpectedly opposite sign. Untabulated analyses reveal that when we exclude $M/B \times Leverage$ from the model, M/B is positive and significant as expected, consistent with hedging increasing with a firm's growth opportunities; in addition, *Leverage* becomes positive and significant as expected. The higher the level of institutional ownership, the more likely a firm is to hedge, consistent with managers responding to external pressures for predictable earnings by hedging oil price risk. The significantly positive *Firm size* coefficient is consistent with the importance of economies of scale in implementing a hedging program. Also as expected, the greater the basis risk due to international production, the lower the probability that oil and gas producers hedge.¹⁴ Finally, the greater the proportion of sales from oil and gas production (i.e., the less diversified the operations), the higher the probability of hedging.

The two right-hand sets of columns in Table 3 report the results of the simultaneous equations estimation. Reported t-statistics are White (1980) adjusted. The coefficient on *PredAAC smoothing ratio* in the *Hedging ratio* regression is not significant ($t = 0.83$). On the other hand, the coefficient on *PredHedging ratio* in the *AAC smoothing ratio* regression is reliably negative, as predicted ($t = -2.41$). Thus, after controlling for other determinants of AAC smoothing, the more firms hedge oil price risk with derivatives, the lower the level of smoothing with AACs. The Hausman (1978) tests for simultaneity (bottom row of Table 3) are consistent with these results: We can reject the null hypothesis of no simultaneity

¹⁴ *Intl production* might proxy for more than basis risk exposure. Firms with overseas oil and gas production may be subject to foreign exchange risk as well as basis risk. Although oil price is denominated in U.S. dollars worldwide, thus protecting revenues from foreign exchange risk, the cost side could be exposed. Thus, firms with international production may have an incentive to hedge foreign exchange risk, and our dummy variable could proxy for that. The observed negative coefficient on *Intl production* in the hedge/no hedge regression would then capture both the impediment to hedging oil price risk due to basis risk and substitution between foreign exchange hedging and oil price hedging.

TABLE 3
Results for Estimation of the Binomial Probit Model of the Decision to Hedge and the Subsequent Two-Stage Least Squares Estimation of the Extent of Hedging and the Extent of Smoothing with Abnormal Accruals

Independent Variables	Dependent Variables					
	Hedgers			AAC Smoothing Ratio		
	Expected Sign	Coeff.	χ^2 Value	Expected Sign	Coeff.	t-statistic
Intercept	?	-1.87	5.24**	?	2.28	0.80
Endogenous variables:						
PredAAC smoothing ratio				-	0.17	0.83
PredHedging ratio						
Common explanatory variables:						
Explrisk	+	0.33	1.79*	+	-0.37	-0.43
FullCost	+/-	-0.00	0.00	+/-	-0.32	-0.57
Leverage	+	-0.88	1.26	+	-0.67	-0.90
M/B	+	-0.68	16.31	+	-0.07	-0.32
M/B $\times$ Leverage	+	0.78	12.51***	+	0.07	0.25
Tax	+	-0.35	0.89	+	0.32	1.91**
MgrIOwn	+/-	0.01	1.67	+/-	0.01	2.43***
Stock options	-	-0.69	0.42	-	-2.63	-1.43*
InstitIOwn	+/-	0.02	9.96***	+/-	-0.00	-2.47***
Firm size	+	0.31	10.74***	+	0.15	0.65
Variables for hedging equations:						
σ_{RET}	-	-2.97	1.37	-	0.48	1.20
Production exposed	-	-0.00	0.00	-	-0.01	-1.30*
Intl production	-	-0.81	9.56***	-	-0.21	-0.59
O&G Production	+	0.01	2.48*	+	0.01	1.69**
Cash		1.07	0.98	-	-1.94	-0.89

(Continued on next page)

TABLE 3 (Continued)

Independent Variables	Dependent Variables							
	Hedgers		Hedging Ratio		AAC Smoothing Ratio		t-statistic	t-statistic
	Expected Sign	Coeff.	χ^2 Value	Expected Sign	Coeff.	Expected Sign	Coeff.	
Variables for AAC smoothing equations:								
<i>DivPayout</i>						+	0.06	2.45**
<i>MarkToMarket</i>						+	-1.71	-1.19
<i>InvMills</i>				?	-0.04	?	6.27	3.12**
Number of observations			236					103
OLS adjusted R ²			30.22%					
Second-stage OLS adjusted R ²								15.85%
Hausman simultaneity test:								
p-value							0.46	0.00

*, **, **** Indicate p-values less than or equal to 0.10, 0.05, 0.01, respectively. p-values are one-tailed if the coefficient is significant in the hypothesized direction, and two-tailed otherwise; t-statistics are White (1980)-adjusted in the *Hedging ratio* and *AAC smoothing ratio* regressions. An OLS adjusted R² is provided for the probit regression for descriptive purposes. Coefficients on year-dummies are not reported.

PredAAC smoothing ratio and *PredHedging ratio* are predicted values of the respective endogenous variables derived from the first-stage of two-stage least squares estimation. *InvMills* is the inverse Mills ratio extracted from the *Hedgers* regression. We assume the decision to hedge is independent of the extent of smoothing with abnormal accruals.

See Table 1 for other variable definitions.

for the *AAC smoothing ratio* regression ($p = 0.00$) but not for the *Hedging ratio* regression. Collectively, these results suggest that oil and gas producers determine the extent of hedging independently of their decisions about smoothing with AACs, but that the extent to which they smooth with AACs is inversely related to the amount they hedged, after controlling for other determinants of smoothing.

Our inferences differ from those in Barton (2001), who finds evidence of simultaneity and substitution between hedging and AACs in both his derivatives and accruals regressions. However, in a sensitivity test, Barton (2001, 21) detects some evidence of sequentiality, consistent with our inference that managers first determine the extent of hedging with derivatives, and then manage residual volatility by trading off smoothing with AACs against hedging. Barton's (2001) sample differs from ours; it covers the period 1994–1996, it is a much larger and broader sample that includes large firms from many industries, 72 percent of his sample use derivatives for foreign exchange and interest rate hedging, and he does not consider commodity derivatives. Possible reasons why Barton (2001) found that the extent of abnormal accruals affects hedging while we did not include the following: (1) because of his broad-based sample spanning many industries, Barton's AAC and derivatives variables are likely noisier than our industry-specific measures, and measurement error is an alternative explanation for his result (see Barton 2001, 3–4, 8, 21; Greene 1993, sections 9.5.5 and 20.5); (2) our smaller sample may provide less powerful tests; and (3) non-oil and gas firms that face interest rate risk associated with long-term debt likely take hedging positions less frequently than firms hedging oil price risk, because derivatives that hedge such interest rate risk typically are in place for longer periods of time than are commodity derivatives that hedge oil price risk. If managers of non-oil and gas firms hedge less frequently and therefore use accruals more frequently to manage residual volatility, then these managers may be more likely to consider their accrual decisions in taking derivative positions. We leave the resolution of the difference in the two studies' inferences for future research.

In the *Hedging ratio* regression, several control variables are significant. There tends to be more extensive hedging in years when firms are profitable and have NOL carryforwards ($t = 1.91$), consistent with Graham and Smith (1999). Managers holding higher percentages of their firms' outstanding shares hedge more ($t = 2.43$), perhaps because they are risk-averse and either unable to diversify firm-specific risks or wish to make their firms appear less risky. There is weak support for less hedging the more stock options that managers hold ($t = -1.43$), whereas there is more extensive hedging the lower the level of institutional holdings ($t = -2.47$). To the extent that lower institutional ownership proxies for less external monitoring, managers may hedge to help external parties gauge profitability due to managerial performance.¹⁵ There also tends to be less hedging the more production exposed to basis risk ($t = -1.30$), and more hedging in firm-years with less diversified operations ($t = 1.69$), consistent with Haushalter (2000). Also consistent with Haushalter (2000), *Firm size* is significant in the hedge/no hedge regression, but not in the extent of hedging regression.

Turning to the *AAC smoothing ratio* regression, in addition to the negative coefficient on *PredHedging ratio*, there are several other significant determinants of AAC smoothing. Full cost firms use AACs to smooth earnings more extensively than do successful efforts firms ($t = 2.07$), suggesting that unidentified factors beyond those controlled here prompt

¹⁵ The significantly negative coefficient on *InstitOwn* is likely not due to collinearity with firm size. In untabulated results using firm-years with hedging, we find significantly lower mean and median *InstitOwn* in firm-years with more extensive hedging.

full cost firms to smooth earnings even more, beyond the smoothing effects of the full cost method.¹⁶ There is more smoothing with AACs in firm-years with larger growth opportunities ($t = 1.40$) and greater costs of underinvestment ($t = 1.87$). Dropping $M/B \times Leverage$ from the *Hedging ratio* and *AAC smoothing ratio* equations does not result in *Leverage* becoming significant, in contrast to the *Hedgers* regression results. Consistent with the *Hedging ratio* regression results, the greater the managerial ownership, the more smoothing, in this case with AACs ($t = 3.06$). However, contrary to the results for the *Hedging ratio* regression, higher levels of institutional holdings are associated with greater smoothing with AACs ($t = 1.74$), suggesting that external pressures may induce managers to use AACs to make reported earnings more predictable. Finally, smoothing with AACs increases with *Firm size* ($t = 2.73$) and with higher dividend payout rates ($t = 2.45$).

Somewhat surprisingly, the results suggest that the level of exploration risk is an important determinant only of the decision to hedge. That is, the higher the level of exploration risk, the more likely managers are to hedge, but the level of exploration risk does not play a significant role in managers' decisions about the extent of hedging or smoothing with AACs.

On the whole, the significance of the results for the control variables in both the hedging and smoothing with accruals regressions are mixed and weak. However, the predicted control variables are not the key hypothesized effects. To better isolate the core effects of hedging on smoothing with AACs and vice versa, we opted to control for factors that might possibly be associated with hedging or smoothing with AACs, even if the rationale for including a control variable was not beyond question. Moreover, we based our predictions for the control variables on expected unconditional relations between the control variables and the extent of smoothing with AACs or hedging. The incremental effect of each control variable, conditional on the inclusion of all other variables in the regressions, may not be the same as the predicted unconditional relation. Our small sample size also likely contributed to some of the control variables' insignificant results.

We tested the null hypothesis that the proportion of significant coefficients in each of the regressions in Table 3 is not greater than the proportion that would be expected by chance. Specifically, we treated the *Hedgers*, *Hedging ratio*, and *AAC smoothing ratio* regressions separately, related the number of coefficients that are significant in the predicted direction to the total number of coefficients with directional predictions, and then used a binomial test of the difference in proportions to investigate whether significantly more than 10 percent of the coefficients in each equation are significant in the predicted direction.¹⁷ In untabulated results, we can reject the null hypothesis at the 0.03 level or better for each of the three regressions. This analysis supports the conclusion that the reported results are not likely a chance occurrence, although it is clear that we need further research to identify more definitively the determinants of income smoothing using AACs or hedging.

Finally, in the *AAC smoothing ratio* regression, the coefficient on the inverse Mills ratio (the selectivity term that measures the covariance between the decisions about whether to hedge and the extent of smoothing with AACs) is positive and significant ($t = 3.12$). Hence, firms that opt to hedge are more likely to smooth with AACs (Heckman 1979; Shehata 1991). Coupled with our primary results, this suggests that although firms that hedge are

¹⁶ Incorporating *FullCost* in the modified Jones model should mitigate any mechanical tendency of full cost to induce larger *AAC smoothing ratio* values.

¹⁷ The test statistic is $(\hat{p} - p_0) / \sqrt{p_0(1 - p_0)/n}$, where $\hat{p}$ is the observed proportion of significant coefficients, p_0 is the null hypothesis proportion (0.10), and n is the total number of coefficients with directional predictions (excluding the intercept and the inverse Mills ratio). We evaluate the significance of the test statistic using a t distribution.

more likely to smooth with AACs, the extent to which they use AACs to smooth earnings is inversely related to the extent they hedge, consistent with trade-offs at the margin by managers using both smoothing tools.¹⁸

Evidence on the Sequential Process

The negative coefficient on the hedging ratio in the *AAC smoothing ratio* regression and the insignificant coefficient on the smoothing ratio in the *Hedging ratio* regression in Table 3 are consistent with managers first deciding whether and how much to hedge, and then substituting between abnormal accruals and hedging to manage residual earnings volatility. If this sequential decision process is valid, then we would expect more of the abnormal accrual vs. hedging trade-off to occur in the fiscal fourth quarter than in the first three fiscal quarters. By the fourth quarter, managers have more accurate information about likely residual earnings volatility, and can make AAC decisions accordingly. Hence, we focus on the effects of the fourth quarter vis-à-vis the first three fiscal quarters to provide more direct evidence on the sequential process.

We recompute the dependent variable in the *AAC smoothing ratio* equation using data from fiscal quarters 1–3 and also obtain a new predicted AAC smoothing variable for the *Hedging ratio* equation; all other variables remain unchanged. The mean *AAC smoothing ratio* before deleting the fourth quarter data is 4.20, as reported in Panel A of Table 2. However, the mean ratio increases to 4.85 when we delete fourth-quarter data, and the difference in means is significant at the 0.07 level. Thus, there is more smoothing in the first three quarters than in the fourth quarter. This is consistent with managers making fiscal fourth quarter adjustments that “settle up” errors in interim quarters’ accruals (Collins et al. 1984; Mendenhall and Nichols 1988; see also Palepu 1988) and also with more non-recurring transactions occurring in the fourth quarter (Elliott and Shaw 1988).

We next test whether our substitution hypothesis holds after deleting fiscal fourth quarter data to compute the AAC smoothing variables. We re-estimate Equations (1)–(3) and obtain the following results (see Panel A of Table 4): (1) In the *Hedging ratio* equation, both the coefficient on *PredAAC smoothing ratio* and the Hausman simultaneity test are insignificant, as in Table 3, but (2) in the *AAC smoothing ratio* equation, the negative relation between the extent of smoothing with AACs and hedging becomes insignificant, as does the Hausman test. Thus, there is no evidence that managers substitute between the extent of hedging and smoothing with AACs in the first three quarters of the year. Consequently, this substitution appears to arise largely in the fourth quarter. This pattern of evidence is consistent with a sequential decision-making process where managers first hedge and then, mostly in the fourth quarter, trade off the use of AACs and hedging with derivatives to manage residual volatility in income.

We extend the analysis by deleting data from the first quarter instead of the fourth quarter. If the sequential process is valid, then we expect that by including fourth-quarter data the coefficient on *PredHedging ratio* will again be significantly negative in the *AAC smoothing ratio* regression, as in Table 3, and the Hausman simultaneity test will also be significant. The results confirm this expectation: the coefficient on the hedging ratio is

¹⁸ We re-estimated the *Hedging ratio* and *AAC smoothing ratio* regressions separately using OLS, and the results are similar to those reported in Table 3. In addition, we tested for serial correlation in the residuals because some firms contribute more than one observation to our sample. For the *AAC smoothing ratio* regression, the Durbin-Watson statistic is 2.06, indicating no first-order serial correlation. For the *Hedging ratio* regression, Durbin-Watson = 1.29 ($p = 0.05$). Understatement of the standard error from the indicated serial correlation is not a serious concern, however, because the coefficient on *AAC smoothing ratio* is insignificant.

TABLE 4

Two-Stage Least Squares Estimation of the Extent of Hedging and the Extent of Smoothing with Abnormal Accruals after Excluding One Fiscal Quarter's Data in Calculating AAC Smoothing Ratio and PredAAC Smoothing Ratio, Leaving All Other Variables Unchanged

Panel A: Excluding Data from the Fiscal Fourth Quarter in Computing the AAC Smoothing Ratio^a

<i>Independent variables</i>	<i>Dependent Variables</i>					
	<i>Hedging Ratio</i>			<i>AAC Smoothing Ratio</i>		
	<i>Expected Sign</i>	<i>Coeff.</i>	<i>t-statistic</i>	<i>Expected Sign</i>	<i>Coeff.</i>	<i>t-statistic</i>
Intercept	?	1.42	0.90	?	-2.45	-0.47
Endogenous variables:						
<i>PredAAC smoothing ratio</i>	-	0.11	1.05			
<i>PredHedging ratio</i>				-	-5.40	-0.90
Common explanatory variables:						
<i>Explrisk</i>	+	0.81	1.13	+	-2.72	-0.59
<i>FullCost</i>	+/-	-0.18	-0.56	+/-	2.17	1.75*
<i>Leverage</i>	+	0.76	1.32	+	-6.04	-1.37
<i>M/B</i>	+	0.07	0.31	+	-1.07	-1.23
<i>M/B × Leverage</i>	+	-0.07	-0.29	+	1.11	1.09
<i>Tax</i>	+	0.49	1.89**	+	-2.46	-1.74
<i>MgrlOwn</i>	+/-	-0.01	-0.56	+/-	0.08	2.81***
<i>Stock options</i>	-	-1.78	-1.89**	-	-7.47	-0.84
<i>InstitOwn</i>	+/-	-0.00	-0.60	+/-	0.05	1.80**
<i>Firm size</i>	+	-0.09	-0.66	+/-	1.03	1.97**
Variables for hedging equations:						
<i>σRET</i>	-	6.47	1.96			
<i>Production exposed</i>	-	-0.01	-1.89**			
<i>Intl production</i>	-	-0.04	-0.16			
<i>O&G Production</i>	+	0.00	0.65			
<i>Cash</i>	-	-0.42	-0.44			
Variables for AAC smoothing equations:						
<i>DivPayout</i>				+	0.07	2.52**
<i>MarkToMarket</i>				+	-4.39	-2.52
<i>InvMills</i>	?	-0.44	-0.51	?	6.23	2.57**
Number of observations			103			103
Second-stage OLS adjusted R ²			10.45%			14.33%
Hausman simultaneity test:						
p-value			0.65			0.45

(Continued on next page)

TABLE 4 (Continued)

Panel B: Excluding Data from the Fiscal First Quarter in Computing the AAC Smoothing Ratio^b

Independent variables	Dependent Variables					
	Hedging Ratio			AAC Smoothing Ratio		
	Expected Sign	Coeff.	t-statistic	Expected Sign	Coeff.	t-statistic
Intercept	?	3.22	0.69	?	-2.27	-0.50
Endogenous variables:						
<i>PredAAC smoothing ratio</i>	-	0.39	0.40			
<i>PredHedging ratio</i>				-	-3.99	-1.83**
Common explanatory variables:						
<i>Explrisk</i>	+	0.62	0.72	+	-1.98	-0.48
<i>FullCost</i>	+/-	0.65	0.33	+/-	2.03	1.83*
<i>Leverage</i>	+	1.41	0.62	+	-2.61	-0.66
<i>M/B</i>	+	-0.36	-0.42	+	-0.44	-0.57
<i>M/B × Leverage</i>	+	0.40	0.06	+	0.54	0.59
<i>Tax</i>	+	0.04	0.95	+	0.50	0.40
<i>MgrlOwn</i>	+/-	-0.01	-0.26	+/-	0.04	1.81*
<i>Stock options</i>	-	-4.31	-1.64*	-	1.25	0.16
<i>InstitOwn</i>	+/-	-0.01	-0.32	+/-	0.06	2.13**
<i>Firm size</i>	+	-0.02	-0.08	+/-	0.41	0.89
Variables for hedging equations:						
<i>σRET</i>	-	6.04	1.84			
<i>Production exposed</i>	-	-0.01	-1.91**			
<i>Intl production</i>	-	-0.73	-0.43			
<i>O&G Production</i>	+	0.02	0.35			
<i>Cash</i>	-	-0.70	-0.81			
Variables for AAC smoothing equations:						
<i>DivPayout</i>				+	0.05	1.87*
<i>MarkToMarket</i>				+	-1.99	-1.28
<i>InvMills</i>	?	0.83	0.26	?	5.38	2.47**
Number of observations			103			103
Second-stage OLS adjusted R ²			8.43%			11.12%
Hausman simultaneity test: p-value			0.47			0.02

*, **, *** Indicate p-values less than or equal to 0.10, 0.05, 0.01, respectively. p-values are one-tailed if the coefficient is significant in the hypothesized direction, and two-tailed otherwise; t-statistics are White (1980)-adjusted in the *Hedging ratio* and *AAC smoothing ratio* regressions.

Coefficients on year-dummies are not reported.

^a *AAC smoothing ratio* is computed using data from fiscal quarters 1-3, and *PredAAC smoothing ratio* is derived from the first stage of two-stage least squares estimated based on the recalculated *AAC smoothing ratio*.

^b *AAC smoothing ratio* is computed using data from fiscal quarters 2-4, and *PredAAC smoothing ratio* is derived from the first stage of two-stage least squares estimation based on the recalculated *AAC smoothing ratio*.

Results for the *Hedgers* regression are unaffected and not repeated here (see Table 3).

All other variables are as defined in Table 3 and Table 1.

significantly negative ($t = -1.83$) and the Hausman test is significant ($p = 0.02$; see Panel B of Table 4). Thus, the primary results remain when we drop first-quarter data, but not when we drop fourth-quarter data.

Control variables with significant coefficients in the AAC *smoothing ratio* regression in Table 3 remain significant in Table 4 except for (1) *M/B* and *M/B* $\times$ *Leverage*, when we drop fourth-quarter data, and (2) *M/B*, *M/B* $\times$ *Leverage*, and *Firm size*, when we drop first-quarter data. In the *Hedging ratio* equation, (1) *Tax*, *Stock options*, and *Production exposed* continue to have significant coefficients, as in Table 3, even after we drop fourth-quarter data, but *MglrOwn*, *InstitOwn*, and *O&G Production* do not, and (2) we obtain similar results when we drop first-quarter data, except *Tax* is no longer significant.

The regressions in Table 4 reflect lower explanatory power than those in Table 3, and we cannot reject the null hypothesis that the proportion of significant coefficients with directional predictions is what we would expect by chance (not shown).¹⁹ The lower explanatory power may be due to time-period misalignment, because we use three fiscal quarters to estimate smoothing-with-AACs variables vs. annual data for all other variables. The lack of availability of quarterly data on derivatives precludes us from computing *Hedging ratio* based on three fiscal quarters' data instead of annual data. The differences in the control variable results between Panels A and B of Table 4 stem from the calculation of the AAC smoothing variables. We compute the AAC smoothing variables using data from quarters 1–3 in Panel A vs. data from quarters 2–4 in Panel B. As noted above, fiscal fourth quarters typically reflect accruals that correct estimation errors from interim quarters (e.g., inventory estimates vs. actual counts) and there may also be more nonrecurring events in the fourth quarter. Thus, computing the AAC smoothing variables using data from the first three quarters should affect the smoothing variables differentially as compared to the case where we compute them using data from quarters 2–4, and, in turn, the relations of these alternatively computed smoothing variables with the control variables in the regressions should also be differentially affected.

Additional Robustness Checks

Finally, we report on a series of tests performed to assess further the sensitivity of the results (details not shown). First, we redefined the denominator of *Hedging ratio* as oil and gas production plus reserves to allow for the possibility that oil price risk extends to reserves. We re-ran the analysis and our inferences are unaffected.

Second, we re-ran the analysis after (1) augmenting Equations (1)–(3) with variables reflecting the interaction of *FullCost* with *PredAAC smoothing ratio*, *Explrisk*, and other control variables that rely on accounting measures and that are thus likely to be systematically affected by the choice of full cost or successful efforts, and (2) including *M/B* or *M/B* $\times$ *Leverage*, but not both, in the models because these variables are highly correlated in the hedge/no hedge regression. Our inferences are unchanged under these alternative model specifications.

Third, Collins and Hribar (2000) identify measurement error in estimating abnormal accruals using changes in balance sheet data instead of cash flow statement data. The problem arises primarily in periods when major acquisitions or divestitures occur. We identified 20 firms for which any sample quarter was affected by a merger or discontinued operations. We eliminated these firms and re-estimated the AAC *smoothing ratio*. The resultant smoothing ratio is insignificantly higher than the full-sample ratio (4.46 vs. 4.20),

¹⁹ We can reject the null hypothesis of chance for the AAC smoothing regressions when we consider both one-tail and two-tail predictions.

and we obtain qualitatively similar results when we re-run the analysis using the re-estimated AAC *smoothing ratio*.

Fourth, we adjust for special items in the AAC *smoothing ratio* calculation. We obtain special items data for all firm-quarters from Compustat (quarterly data item #32), tax-adjust the data (multiplying by $1 - \text{tax rate}$, or 0.65), and purge the resulting loss or gain from cash flows from operations. The mean AAC *smoothing ratio* becomes 4.21 vs. 4.20 before such adjustment, the difference between the two ratios is not significant, and our inferences are unaffected when we re-run our primary analysis using the adjusted AAC *smoothing ratio*.²⁰

V. SUMMARY AND CONCLUSIONS

We examine whether managers use abnormal accrual choices and oil and gas price hedging with derivatives as substitute mechanisms to manage earnings volatility. Our goal is to investigate how accounting decisions affect hedging, and vice versa. Firms engaged in oil and gas exploration and drilling face two kinds of inherent risks that induce volatility in their earnings streams. Firms can hedge oil price risk with derivative instruments, but markets do not exist in which firms can hedge the operational risk of unsuccessful drilling. Firms can, however, use both hedging and abnormal accruals to dampen the impact that oil price and exploration risks have on earnings variability. We find that once managers of oil and gas producing firms decide to hedge, they trade off the extent of smoothing with abnormal accruals against the extent of hedging, at the margin. More specifically, the results suggest that managers of oil and gas producing firms take commodity hedging positions independent of their decisions about abnormal accruals, but then, primarily in the fourth quarter, substitute between abnormal accruals and hedging with derivatives to manage residual volatility.

There are several limitations of our study. First, we have data only on year-end derivative positions and thus cannot gauge derivative activity throughout the year. Second, our smoothing-with-AACs variable captures managerial discretion with measurement error. Third, we have not treated the full cost vs. successful efforts accounting choice as an endogenous element of a firm's overall strategy for managing earnings volatility, although we incorporate the choice when estimating abnormal accruals and also as a determinant of hedging and smoothing with abnormal accruals. Fourth, our conclusion that a sequential hedging-then-abnormal-accruals decision process characterizes oil and gas managers' behavior contrasts with Barton's (2001) overall conclusion of a simultaneous process whereby abnormal accruals affect hedging, as well as hedging affecting abnormal accruals. We are not able to resolve the differences between our inferences and those in Barton (2001), and leave that task for future research. Finally, our results may not generalize to other populations because of our single-industry oil and gas producer setting, and because our data requirements bias the sample toward larger producing firms that are income smoothers.

²⁰ We also examined firms' financial reports for special items, and they typically reflect activities related to oil and gas exploration and production. These include (1) provisions for dismantlement, restoration, and reclamation based on estimates of, for example, environmental clean ups, and (2) provisions for asset impairments, including write-downs of oil and gas properties. There are more frequent and larger-magnitude oil and gas reserve write-downs in firm-years with hedging vs. those with no hedging, but a few extreme impairments reported in October 1995 under SFAS No. 121 (FASB 1995) primarily drive the magnitude difference.

APPENDIX A

**Illustration of the Computation of the Hedging Ratio from SFAS No. 69 Disclosures
and Commodity Derivative Disclosures from the 1995 Form 10-K of Newfield
Exploration Company**

I. SFAS No. 69 Disclosures

**SUPPLEMENTARY FINANCIAL INFORMATION
SUPPLEMENTARY OIL AND GAS DISCLOSURES-UNAUDITED
(paragraphs omitted)**

<i>Proved Developed and Undeveloped Reserves</i>	<i>Oil, Condensate and Natural Gas Liquids (MBbls)</i>	<i>Natural Gas (MMcf)</i>
December 31, 1994	8,610	153,967
Production	(2,071)	(33,719)
Other reserve changes during the year (details suppressed)	<u>3,094</u>	<u>83,332</u>
December 31, 1995	<u>9,633</u>	<u>203,580</u>

Reserve Quantities:

For the year ended December 31, 1995, quantity of proven oil reserves is 9.63 million barrels. Proven gas reserves of 203,580 million cubic feet, converted at the rate of 9,840 cubic feet to a barrel of oil (based on the ratio of average sale prices for oil and gas realized by the firm during 1995) is equivalent to 20.68 million barrels. Hence, the total reserves of the firm are 30.31 million barrels.

II. Derivative Disclosures (Voluntary Disclosures as per SFAS No. 119)

From time to time, the Company has utilized hedging transactions with respect to a portion of its oil and gas production to achieve a more predictable cash flow, as well as to reduce its exposure to price fluctuations.

(paragraphs omitted)

The following is a summary of the Company's gas swap positions as of December 31, 1995 and 1994.

	<i>MMcf</i>	<i>Period</i>	<i>Weighted Average Price (MMBtu)</i>	<i>Weighted Average Price (Mcf)</i>	<i>Fair Market Value</i>
December 31, 1995	15,000	January 1996– September 1996	\$1.72	\$1.83	(\$7.2 million)
December 31, 1994	11,150	January 1995– September 1995	\$1.72	\$1.84	\$1.9 million

Oil. The Company has entered into sales contracts for approximately 421 barrels of oil production per day for the period January 1996 through April 1996, which effectively fixed the Louisiana Light Sweet ("LLS") posted price for such production at \$16.50 per barrel.

Additionally, the Company has entered into a crude oil swap agreement for 1,000 barrels of oil production per day for the period January 1996 through June 1996, which effectively fixed the LLS posted price for such production at \$15.25 per barrel.

III. Our Study's Use of Derivative Disclosures to Compute the Hedge Ratio

1. Notional quantity of oil derivatives-short is the sum of

Oil swaps: 1,000 barrels/day for 180 days	0.182 million barrels
LLS contracts: 421 barrels/day for 120 days	0.050 million barrels
	<u>0.232 million barrels</u>

2. Notional quantity of gas derivatives is 15,000 MMcf or 1.524 million barrels of oil when converted at 9,840 cubic feet to a barrel of oil. Conversion is based on the ratio of the average sale price for oil and gas the firm realized during 1995.
3. Total notional quantity of derivatives: 1.756 million barrels of oil.
4. Production for the year is 5.497 million barrels. The production quantity is computed as 2.071 million barrels of oil (see disclosures in part I above) and 3.426 million barrels equivalent to 33,719 MMcf of gas (converted to oil at the conversion rate used in no. 2, above).
4. The company's *Hedging ratio* for 1995 is the total notional quantity of derivatives scaled by the notional quantity of production: 1.756 million barrels/5.497 million barrels = **31.94 percent**.

APPENDIX B

Derivation and Computation of Sunder's Variance of Cash Flows from Exploration I. Summary of Sunder's (1976) Theoretical Derivation

Consider a firm that drills N exploratory wells each period. The probability of a successful strike (θ) is the same each period.²¹ The nonrecoverable exploration cost of each well is c . Each successful well yields a net operating revenue of x per period for L periods starting the period after drilling takes place.

S_t , the number of wells drilled in period t , is a random variable with a binomial distribution and parameters θ and N . The probability (Pr) that S equals an integer r between zero and N is given by: $Pr(S = r|\theta, N) = [N!/(N-r)!r!] \times [\theta^r(1-\theta)^{N-r}]$. The expected value and variance of S are $N\theta$ and $N\theta(1-\theta)$, respectively.

Revenue generated by drilling efforts is also a random variable. If S wells are successful, then this will yield operating revenues of Sx for L future periods beginning the next period, where x is the net operating cash flow per successful exploratory well per period for L periods. Therefore, the net future cash inflow in period t , X_t , will be: $X_t = -Nc + x(S_{t-1} + S_{t-2} + \dots + S_{t-L})$. The mean of future net cash flows is $-Nc + xLN\theta$ and the variance is as follows:

²¹ Assuming a constant θ implies that θ and N are independent. One can argue that θ depends on experience in previous periods (e.g., learning effects) or that θ is inversely related to N (e.g., firms probably exhaust better drilling prospects first). Because the components of our empirical measure of Sunder's variance can vary each year, we allow for the possibility that θ changes over time.

$$\text{Var}(X) = x^2 \text{LN}\theta(1 - \theta) = \text{PV}^2(1 - \theta)/\text{LN}\theta \quad (\text{B1})$$

where PV is the present value of discoveries from exploration activity.²²

Sunder (1976) does not indicate whether the parameters in Equation (B1) are to be estimated annually or as firm-level averages. We compute Sunder's variance annually because year-to-year differences in oil price levels can affect some of the parameters in Equation (B1). For example, the economic life of reserves and operating cash flows likely increase with oil prices.

II. Computing the Variance of Future Cash Flows from Exploration Activities

The following are excerpts of disclosures by Newfield Exploration Company for 1995 under SFAS No. 69. They represent the present value of discoveries from exploration activity; i.e., the changes in the standardized measure of discounted future net cash flows applicable to proved oil and gas reserves (in thousands of dollars):

	<i>Year ended December 31, 1995</i>
Beginning of the period	\$180,002
Revisions of previous estimates:	
Additions to proved reserves resulting from extensions, discoveries and improved recovery, less related costs ($x\text{LN}\theta - N_c$)	87,760
Other reasons for the change (suppressed here)	<u>8,564</u>
End of period	<u>\$276,326</u>

Based on the company's SFAS No. 69 disclosures for 1995, exploration costs (N_c) are \$32.50 million. Thus, $x\text{LN}\theta$, which equals $(x\text{LN}\theta - N_c) + N_c$, is $\$87.76 + \$32.50$ or \$120.26 million.

We now compute Sunder's variance using Newfield Exploration Company's 1995 data:

(a) *Productive life of a well (L)*

Using numbers drawn from part I of Appendix A, the productive life of a well is:
 $L = [(\text{Beginning oil reserves}/\text{Oil production}) + (\text{Beginning gas reserves}/\text{Gas production})]/2$
 $= [(8,610/2,071) + (153,967/33,719)]/2 = 4.362$ years. (Our inferences are unchanged when we compute age as a weighted average of the ages of oil and gas wells instead of a simple average.)

(b) *Number of exploratory wells drilled (N) and the success rate (θ)*

The following table reports the company's drilling activity for 1995:

²² Sunder's variance is unaffected by the choice of full cost or successful efforts, although it is affected by hedging, which affects cash flows. This likely reflects a second-order effect of the FC/SE choice on cash flows and hence on Sunder's variance. Exploration-related cash flows likely are measured before the impact of hedging, since we have not seen statements by sample firms that they hedge expected production from discoveries or exploration. They are more likely to hedge production from the extant stock of reserves.

	<u>Gross²³</u>	<u>Net</u>
Exploratory wells:		
Productive	10	4.7
Nonproductive	<u>6</u>	<u>3.9</u>
Total	<u>16</u>	<u>8.6</u>

Number of exploratory wells (N) is 8.6 and success rate (θ) is $4.7/8.6 = 54.6$ percent. Thus, 8.6 and 54.6 percent may be viewed as realizations from an underlying firm-specific distribution of N and θ . (If the number of exploratory wells drilled is 0 and/or the success rate is 0 or 1, then we set the variance equal to 0.)

(c) Present value of cash flows per productive period (x)

We compute x by dividing $xLN\theta$, the present value of discoveries from exploration activity, by [Productive life of a well (L) $\times$ Number of exploratory wells drilled (N) $\times$ Success rate (θ)] = $xLN\theta/LN\theta = \$120.26$ million/(4.362 years $\times$ 8.6 wells $\times$ 0.546) = \$5.87 million.

(d) Variance of future cash flows from exploration activity

We compute the variance as $(xLN\theta) \times [x(1 - \theta)] = x^2LN\theta(1 - \theta) = \$120.26$ million $\times$ [\$5.87 million $\times$ (1 - 0.546)] = \$320.49 million. Finally, we divide the variance of future cash flows from exploration activity by the value of year-end reserves, calculated as (quantity of year-end oil reserves $\times$ year-end spot price of oil) + (quantity of year-end gas reserves $\times$ year-end spot price of gas).

²³ Firms frequently conduct exploration activities as part of a consortium. A *gross well* is a well in which the firm owns an interest. A *net well* represents the fractional interest the firm owns in the gross well. We use net wells to compute exploration success.

REFERENCES

- Barth, M., J. Elliott, and M. Finn. 1999. Market rewards associated with patterns of increasing earnings. *Journal of Accounting Research* 37 (2): 387–413.
- Barton, J. 2001. Does the use of financial derivatives affect earnings management decisions? *The Accounting Review* 76 (1): 1–26.
- Beaver, W., P. Kettler, and M. Scholes. 1970. The association between market determined and accounting determined risk measures. *The Accounting Review* 45 (4): 654–682.
- Belsley, D., E. Kuh, and R. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York, NY: Wiley.
- Bernard, V., and D. Skinner. 1996. What motivates managers' choice of discretionary accruals? *Journal of Accounting and Economics* 22 (1-3): 313–325.
- Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics* 11 (2-3): 255–274.
- Collins, D., and P. Hribar. 2000. Errors in estimating accruals: Implications for empirical research. Working paper, The University of Iowa, Iowa City, IA.
- Collins, W., W. Hopwood, and J. McKeown. 1984. The predictability of interim earnings over alternative quarters. *Journal of Accounting Research* 22 (Supplement): 467–479.
- Cragg, J. 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 39 (5): 829–844.
- Dechow, P., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193–225.

- DeFond, M., and C. Park. 1997. Smoothing in anticipation of future earnings. *Journal of Accounting and Economics* 23 (2): 115–139.
- DeMarzo, P., and D. Duffie. 1995. Corporate incentives for hedging and hedge accounting. *Review of Financial Studies* 8 (3): 743–772.
- Elliott, J., and W. Shaw. 1988. Write-offs as accounting procedures to manage perceptions. *Journal of Accounting Research* 26 (Supplement): 91–126.
- Fargher, N., D. G. Haushalter, and G. F. Peters. 1997. Managing product market and exploration risk in the oil and gas industry—hedging and accounting policy choice. Working paper, University of Oregon, Eugene, OR.
- Financial Accounting Standards Board (FASB). 1982. *Disclosures About Oil and Gas Producing Activities*. Statement of Financial Accounting Standards No. 69. Stamford, CT: FASB.
- . 1994. *Disclosure About Derivative Instruments and Fair Value of Financial Instruments*. Statement of Financial Accounting Standards No. 119. Stamford, CT: FASB.
- . 1995. *Accounting for the Impairment of Long-Lived Assets*. Statement of Financial Accounting Standards No. 121. Stamford, CT: FASB.
- Foster, G. 1980. Accounting policy decisions and capital market research. *Journal of Accounting and Economics* 2 (1): 29–62.
- Fox, J. 1997. Learn to play the earnings game (and Wall Street will love you). *Fortune* 135 (March 31): 76–80.
- Froot, K., D. Scharfstein, and J. Stein. 1993. Risk management: Coordinating investment and financing policies. *The Journal of Finance* 48 (5): 1629–1658.
- Geczy, C., B. Minton, and C. Schrand. 1997. Why firms use currency derivatives. *The Journal of Finance* 52 (4): 1323–1354.
- , ———, and ———. 1999. Choices among alternative risk management strategies: Evidence from the natural gas industry. Working paper, University of Pennsylvania, Philadelphia, PA.
- Graham, J., and D. Rogers. 1999. Is corporate hedging consistent with value maximization? An empirical analysis. Working paper, Duke University, Durham, NC.
- , and C. Smith, Jr. 1999. Tax incentives to hedge. *The Journal of Finance* 54 (6): 2241–2262.
- Greene, W. 1993. *Econometric Analysis*. 2nd edition. Englewood Cliffs, NJ: Prentice Hall.
- Guay, W. R. 1999. The impact of derivatives on firm risk: An empirical examination of new derivative users. *Journal of Accounting and Economics* 26 (1-3): 319–351.
- Han, J., and S-W. Wang. 1998. Political costs and earnings management of oil companies during the 1990 Persian Gulf crisis. *The Accounting Review* 73 (1): 103–117.
- Hand, J. 1989. Did firms undertake debt-equity swaps for an accounting paper profit or true financial gain? *The Accounting Review* 64 (4): 587–623.
- Haushalter, G. D. 2000. Financing policy, basis risk, and corporate hedging: Evidence from oil and gas producers. *The Journal of Finance* 55 (1): 107–152.
- Hausman, J. 1978. Specification tests in econometrics. *Econometrica* 46 (6): 1251–1271.
- Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153–161.
- Holthausen, R., D. Larcker, and R. Sloan. 1995. Business unit innovation and the structure of executive compensation. *Journal of Accounting and Economics* 19 (2-3): 279–313.
- Hughes, K. E., II. 2000. The value relevance of nonfinancial measures of air pollution in the electric utility industry. *The Accounting Review* 75 (2): 209–228.
- Hunt, A., S. Moyer, and T. Shevlin. 1996. Managing interacting accounting measures to meet multiple objectives: A study of LIFO firms. *Journal of Accounting and Economics* 21 (3): 339–374.
- , ———, and ———. 1997. Earnings volatility, earnings management, and equity value. Working paper, University of Washington, Seattle, WA.
- Kennedy, P. 1998. *A Guide to Econometrics*. 4th edition. Cambridge, MA: MIT Press.
- Lambert, R. 1984. Income smoothing as rational equilibrium behavior. *The Accounting Review* 59 (4): 604–618.
- Levitt, A. 1998. The numbers game. Remarks delivered at the NYU Center for Law and Business, New York, NY, September 28.

- Loomis, C. 1999. Lies, damned lies, and managed earnings: The crackdown is here. *Fortune* 140 (August 2): 74–78, 82, 84, 88, 90, 92.
- Malmquist, D. 1990. Efficient contracting and the choice of accounting method in the oil and gas industry. *Journal of Accounting and Economics* 12 (1-3): 173–205.
- Mendenhall, R., and W. Nichols. 1988. Bad news and differential market reactions to announcements of earlier-quarter versus fourth-quarter earnings. *Journal of Accounting Research* 26 (Supplement): 63–86.
- Mian, S. 1996. Evidence on corporate hedging policy. *Journal of Financial and Quantitative Analysis* 31 (3): 419–439.
- Minton, B., and C. Schrand. 1999. The impact of cash flow volatility on discretionary investment and the costs of debt and equity financing. *Journal of Financial Economics* 54 (3): 423–460.
- Palepu, K. 1988. Discussion of bad news and differential market reactions to announcements of earlier-quarter versus fourth-quarter earnings. *Journal of Accounting Research* 26 (Supplement): 87–90.
- Petersen, M. A., and S. R. Thiagarajan. 2000. Risk measurement and hedging: With and without derivatives. *Financial Management* 29 (4): 5–30.
- Rajgopal, S. 1999. Early evidence on the informativeness of the SEC's market risk disclosures: The case of commodity price risk exposure of oil and gas producers. *The Accounting Review* 74 (3): 251–280.
- , and T. Shevlin. 2002. Empirical evidence on the relation between stock option compensation and risk taking. *Journal of Accounting and Economics* (forthcoming).
- Ronen, J., and S. Sadan. 1981. *Smoothing Income Numbers: Objectives and Implications*. Reading, MA: Addison-Wesley.
- Schipper, K. 1989. Commentary on earnings management. *Accounting Horizons* 3 (4): 91–102.
- Schrand, C. 1994. An evaluation of the effects of SFAS No. 80 on interest rate risk management in the savings and loan industry. Working paper, University of Pennsylvania, Philadelphia, PA.
- , and J. Elliott. 1998. Commentary on risk and financial reporting: A summary of the discussion at the 1997 AAA/FASB conference. *Accounting Horizons* 12 (3): 271–282.
- , and H. Unal. 1998. Hedging and coordinated risk management: Evidence from thrift conversions. *The Journal of Finance* 53 (3): 979–1013.
- Shehata, M. 1991. Self-selection bias and the economic consequences of accounting regulation: An application of two-stage switching regression to SFAS No. 2. *The Accounting Review* 66 (4): 768–787.
- Smith, C., and R. Stulz. 1985. The determinants of firms' hedging policies. *Journal of Financial and Quantitative Analysis* 20 (4): 391–405.
- , and J. Warner. 1979. On financial contracting: An analysis of bond covenants. *Journal of Financial Economics* 7 (2): 117–162.
- Subramanyam, K. 1996. The pricing of discretionary accruals. *Journal of Accounting and Economics* 22 (1-3): 249–281.
- Sunder, S. 1976. Properties of accounting numbers under full costing and successful-efforts costing in the petroleum industry. *The Accounting Review* 51 (1): 1–18.
- Trueman, B., and S. Titman. 1988. An explanation for accounting smoothing. *Journal of Accounting Research* 26 (Supplement): 127–139.
- Tufano, P. 1996. Who manages risk? An empirical examination of risk management practices in the gold mining industry. *The Journal of Finance* 51 (4): 1097–1137.
- Warfield, T., J. Wild, and K. Wild. 1995. Managerial ownership, accounting choices, and informativeness of earnings. *Journal of Accounting and Economics* 20 (1): 61–91.
- White, H. 1980. A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4): 817–838.
- Wong, M. H. F. 2000. The association between SFAS No. 119 derivatives disclosures and the foreign exchange exposure of manufacturing firms. *Journal of Accounting Research* 38 (2): 387–417.

Ex Ante and *Ex Post* Discretion over Arm's Length Transfer Prices

Michael J. Smith
Duke University

ABSTRACT: Treasury Regulations typically require a firm to use its manufacturing cost as one input in calculating the transfer price to be used for tax purposes. Because the characteristics of the firm's capital assets determine the realized manufacturing cost, the firm has an incentive to distort its investment choices in order to receive a more favorable tax transfer price. This represents an *ex ante* form of shifting income across divisions. Practical difficulties in enforcing the regulations provide the firm with another way to shift income. In particular, regulators often can identify only a range of acceptable transfer prices. As a result, the firm will not use the true "arm's length" transfer price, but will instead choose a transfer price that is more favorable for tax purposes. I find that the opportunity to shift income *ex post* in this manner may reduce the value of investment-based *ex ante* income shifting. That is, if the firm has more discretion over the *ex post* transfer price, then it may invest more efficiently *ex ante*. Although it is commonly perceived that *ex post* income shifting is purely deleterious, the analysis shows that it may have benefits.

Keywords: *transfer price; tax; multinational firm; regulation.*

I. INTRODUCTION

This paper analyzes the efficiency implications of multinationals' ability to influence, directly or indirectly, the transfer prices they use to allocate profits across divisions for tax purposes. The Internal Revenue Code §482 "arm's length" regulations link the calculation of a multinational's transfer price to a comparable transaction between uncontrolled (independent) parties. All other things equal, a multinational firm wants to set the transfer price to allocate all profit to the division in the lower tax jurisdiction. Uncontrolled parties presumably have no such tax incentive when they negotiate a price with each other. As a result, an uncontrolled transaction price represents a "fair" division of profit across divisions, at least according to the logic underlying the arm's length philosophy.

This paper is based on an essay from my dissertation at Stanford University. I thank the members of my committee, David Baron, George Foster, and Richard Lambert (chair). I also appreciate the comments and suggestions of Jack Hughes, Richard Sansing (associate editor), and two anonymous referees. I gratefully acknowledge the financial support of the Deloitte & Touche Foundation.

*Submitted April 2000
Accepted July 2001*

Under the regulations, the firm derives a fair transfer price for tax purposes by observing (or inferring) the price of an actual (or hypothetical) uncontrolled sale of a good comparable to the good transferred between the multinational's divisions.

Despite these regulations, a firm can still shift income in two ways. First, because it is difficult to identify comparable, uncontrolled transactions, the tax auditor often can identify only a range of acceptable transfer prices. Hence, the auditor cannot always detect a deviation from the "true" arm's length price (*ex post* discretion). Second, because the firm's investment choices may affect the definition of comparability, the firm can manipulate the transfer price *ex ante* by distorting investment.¹ The common perception is that the lax enforcement and/or ambiguity of existing regulations give firms too much *ex post* discretion over the tax transfer price (e.g., Graham 1992; Judis 1993). A comprehensive analysis of *ex post* discretion, however, must also consider its effect on *ex ante* investment.

In the model, two divisions in different tax jurisdictions produce and sell a product. The upstream (selling) division begins the manufacturing process. The downstream (buying) division completes the manufacturing and sells the product outside the group. The firm makes capital investments in each division at time 0. The investments are specializations that are valuable only if the divisions trade with each other, and these investments determine each division's variable cost in subsequent periods. There is no agency conflict in the firm.

I analyze two different types of arm's length price: comparable profit method (CPM) price and comparable uncontrolled price (CUP). Under the CPM, the transfer price must produce a rate of return for the division that is equal to the rate of return of a similar independent firm. Under the CUP method, the transfer price must be equal to that of a similar or identical transaction between independent firms. I analyze both methods, deriving the optimal level of capital in both divisions given different tax rates and different levels of *ex post* discretion.

The study's main result is that the firm's opportunistic selection of the *ex post* transfer price may have efficiency benefits. That is, allowing the firm discretion to deviate from the "true" arm's length price in subsequent periods may mitigate its incentive to choose less efficient investments at time 0. Although in some cases *ex post* income shifting exacerbates investment distortions, in others, it alleviates it. A key distinction between the types of income shifting underlies the result that *ex post* discretion may have efficiency benefits. *Ex ante* discretion over the transfer price affects the absolute amount of income the firm divides between jurisdictions, but *ex post* discretion affects only the relative allocations. In other words, *ex ante* shifting determines the size of the pie; *ex post* shifting determines the relative sizes of the two pieces. Choosing more efficient inputs may result in a regulatory transfer price that is less favorable for tax purposes, but that also produces higher pre-tax profits. From the firm's perspective, the higher pre-tax profits may dominate the higher average tax rate. The results suggest that in some cases an increase in *ex post* discretion may lead to increases in the tax revenues of both jurisdictions. In any case, the interplay between the types of shifting implies that *ex post* shifting has an efficiency, as well as a distributional, role.

The paper builds on the research of Halperin and Srinidhi (1987). They examine the implications of different methods of calculating the CUP transfer price for manufactured goods on single-period resource allocation decisions. Halperin and Srinidhi (1996) perform a similar analysis for the arm's length regulations for intangibles, using both the CUP method and the CPM. In both papers, the authors (1) suppress intrafirm incentive problems,

¹ I define *ex ante* and *ex post* relative to the realization of cost and revenue.

(2) implicitly assume that regulators can perfectly enforce arm's length rules, which eliminates *ex post* shifting, and (3) characterize the *ex ante* resource allocation distortions the regulations induce.

For both the CPM and the CUP method, I also analyze a pure *ex ante* shifting setting such as Halperin and Srinidhi (1987, 1996). In the CUP setting, the production distortion depends only on the tax rate differential. With the CPM, the distortion also depends on the after-tax comparable rate of return, the firm's discount rate, and its depreciation policy. This result holds because increasing investment has conflicting effects on a division's rate of return. By reducing variable cost, higher investment increases profits (the numerator in the rate of return calculation). However, it also increases the capital base (the denominator in the calculation). Therefore, tax rate differences alone do not determine the direction of the investment distortions. Because the CPM results depend on firm-specific factors, analysis of the CUP method yields the most general conclusions.

The study's main contribution is that it is the first to consider *ex ante* and *ex post* shifting simultaneously. Analysis of the feedback between the types of discretion is important because the results undermine the notion that *ex post* discretion benefits only the firm and the lower tax rate jurisdiction. To the extent that regulators wish to promote efficiency as well as a fair distribution of income, the improved efficiency that may accompany higher *ex post* discretion benefits them. The improved efficiency also mitigates, and potentially fully offsets, the reduction in tax receipts suffered by the higher tax rate jurisdiction.

Although Samuelson (1982) and Kant (1988) model similar issues, neither paper examines the interaction between the two types of shifting. Both papers consider a multinational firm that produces a product domestically and sells it both domestically and abroad. Neither model requires that the firm transfer the product at the domestic selling price, an assumption that violates current institutional practice and limits the applicability of the results.

In Samuelson (1982), the transfer price must lie between the marginal cost of production and the domestic selling price. The range of defensible prices represents *ex post* discretion. He shows that the firm distorts production decisions to influence the range of defensible prices, which is *ex ante* shifting. However, he analyzes the interaction in only one direction. He shows how changes in inputs affect the *ex post* range of prices, but does not show how changes in the *ex post* range affect the firm's *ex ante* incentive to distort inputs.

Kant (1988) models the effect of varying *ex post* discretion, but omits *ex ante* shifting. Because the firm's input decisions cannot affect the exogenously specified arm's length price in his model, there is no *ex ante* shifting. Although the firm can choose an arbitrary transfer price, the probability of regulatory intervention increases as the selected transfer price deviates from the arm's length price. The firm balances income shifting against penalty avoidance. Kant (1988) models *ex post* discretion by varying the parameters of the function governing the probability of regulatory intervention, but does not address the interaction between the two modes of income shifting.

The paper is also related to Harris and Sansing (1998) and Sansing (1999). These papers extend the tax transfer pricing literature by explicitly modeling the economic environment that permits the coexistence of both vertically integrated and independent firms. As Sansing (1999) observes, "only transactions for which neither organizational structure dominates will feature both the need for a transfer price for the vertically integrated group and a pair of comparable independent firms from which to derive a transfer price." Harris and Sansing (1998) model the CUP method. They find that it allocates a disproportionate (compared to

independent firms) amount of income to the manufacturer. They also show that it distorts production decisions if tax rate differences exist, and that these distortions may in turn alter the choice of organizational structure. Sansing (1999) abstracts away from efficiency effects and instead examines the relative distributional properties of the CUP method and the CPM. He finds that the CUP method allocates relatively less taxable income to the domestic parent than the CPM because the CPM directly compensates the parent for its relatively larger share of the costs of the relationship-specific investments.

The economic decisions of uncontrolled firms provide the benchmark for arm's length prices. Only by explicitly modeling these decisions, then, can Harris and Sansing (1998) and Sansing (1999) compare the CUP and CPM profit allocations to each other and to the allocation between independent firms. In contrast, I make no comparisons across the CUP method and CPM, restricting my analysis within categories. For expositional clarity, I assume the existence of economic conditions supporting the coexistence of both vertically integrated and independent firms rather than deriving the conditions explicitly. In other words, I take the return of the comparable uncontrolled firm (CPM) and the divisions' external opportunities (CUP method) as given and fixed.² Furthermore, I suppress agency considerations that might favor one organizational form over the other.

The results of the paper are potentially of interest to accounting and tax researchers, tax auditors, regulators, and policymakers. Earlier research in accounting and economics demonstrates that literal adherence to the Treasury Regulations induces firms to distort resource allocations. Also, the business press suggests that lax enforcement of regulations and/or regulations too ambiguous to enforce strictly lead to substantial underpayment of taxes. For example, transfer pricing was briefly an issue in the 1992 presidential campaign, with Bill Clinton claiming that improved IRS enforcement of existing regulation could net \$45 billion in tax revenues (see Graham 1992; Judis 1993). Although the magnitude of the alleged underpayment may be politically motivated hyperbole, the claim illustrates the common perception about the urgency of reducing firms' *ex post* choice over tax transfer prices.

This study shows that, contrary to popular perception, *ex post* discretion may have efficiency benefits. Although still acting opportunistically, firms may react to increased latitude in choosing transfer prices by producing more efficiently. These results can help policymakers understand the implications of transfer pricing regulations. They could also help tax auditors decide how to allocate scarce resources. For example, *ex post* shifting exacerbates *ex ante* distortions under the CUP method if the downstream tax rate is higher, but alleviates them if the upstream rate is higher. Tax authorities would derive more benefits from stricter enforcement in the former setting. Finally, the study is of interest to management accountants who design and implement transfer pricing policies in that it enriches the understanding of the institutional context and implications of transfer pricing policy.

Section II introduces the model. Sections III and IV present the analysis. Section V discusses the results. Finally, Section VI summarizes the paper.

II. THE MODEL

Upstream and downstream affiliates of a controlled group, one affiliate incorporated in the U.S., the other incorporated in another country, combine in the manufacture and sale of a product. The upstream division begins production and transfers the intermediate good to the downstream division. The downstream division finishes production and sells the product outside the group at price p .

² Harris and Sansing (1998) argue that if the decisions of the vertically integrated firm affect the independent transactions, the transactions are not appropriate arm's length benchmarks anyway.

At time 0, the firm makes variable cost-reducing capital investments of k_u and k_d , where the subscripts u and d designate upstream and downstream, respectively. The capital investments allow the divisions to produce a total of q units over N years at variable costs $v_u(k_u)/\text{unit}$ and $v_d(k_d)/\text{unit}$, respectively, with $v'_u(\cdot)$ and $v'_d(\cdot) < 0$, and $v''_u(\cdot)$ and $v''_d(\cdot) > 0$. I assume that neither division's investment affects the other's cost: $\partial v_u(\cdot)/\partial k_d = \partial v_d(\cdot)/\partial k_u = 0$. Without loss of generality I set $q = 1$. I will subscript the time period by $n \in \{1, 2, \dots, N\}$. Let the proportion of the total units produced and sold in period n be γ_n . That is, in year n , the firm produces γ_n units, where $\sum_{n=1}^N \gamma_n = 1$. The firm's depreciation factor in period n is δ_n , with $\delta_n > 0$ and $\sum_{n=1}^N \delta_n = 1$. Let $\lambda_n k_j$ be the book value of asset j at the beginning of period n . Therefore, $\lambda_n = 1 - \sum_{j=1}^{n-1} \delta_j$. Table 1 in Appendix A provides a glossary of notation.

The investment is relationship-specific and has value only if the divisions trade internally. For example, the capital investments could involve specialization of the intermediate product that has no value unless the other manufacturing facility conforms to the specialization. The work of Williamson (1985) motivates this assumption. By organizing as a single firm, economic entities may avoid the "hold-up" problem associated with relationship-specific investment. Hence, it is a sensible setting in which to examine transfer pricing.

I assume that the firm is a price-taker and that the price p is an intertemporal constant.³ In period n , the investments yield pre-tax accounting income of $p - v_u(k_u) - v_d(k_d) - \delta_n(k_u + k_d)$. The divisional profits depend on the transfer price, T_n . The upstream division's pre-tax income in period n is $\gamma_n[T_n - v_u(k_u)] - \delta_n k_u$; the downstream division's is $\gamma_n[p - T_n - v_d(k_d)] - \delta_n k_d$. The upstream division's income is taxed at t_u and the downstream's at t_d . Let $\tau_j = 1 - t_j$. The firm's cash flow in period n is:

$$CF_n = \gamma_n[\tau_u(T_n - v_u(k_u)) + \tau_d(p - T_n - v_d(k_d))] + t_u \delta_n k_u + t_d \delta_n k_d$$

equivalent to:

$$CF_n = \gamma_n[\tau_d(p - v_d(k_d)) - \tau_u v_u(k_u)] + t_u \delta_n k_u + t_d \delta_n k_d + \gamma_n(\tau_u - \tau_d)T_n.$$

Let the firm's discount rate be i and $D_n = (1 + i)^{-n}$. The firm selects k_u and k_d to maximize the net present value of the cash flows.⁴ The net present value over the life of the investment is:

$$\begin{aligned} NPV = & -(k_u + k_d) + \sum_{n=1}^N D_n \gamma_n [\tau_d(p - v_d(k_d)) - \tau_u v_u(k_u)] \\ & + \sum_{n=1}^N D_n \delta_n (t_u k_u + t_d k_d) + (\tau_u - \tau_d) \sum_{n=1}^N D_n \gamma_n T_n. \end{aligned} \quad (1)$$

The first term represents the time 0 investments, the second term is the present value of the after-tax margin, the third term is the present value of the depreciation tax benefit, and the last term is the present value of the after-tax net transfer payments. As a benchmark, consider the case in which the transfer price is independent of the level of capital investment. In this case, the last term in the objective function in Equation (1) is irrelevant.

³ Since k_u and k_d do not affect p , all of the results hold with stochastic p .

⁴ Survey evidence by Graham and Harvey (2001) indicates that firms are as likely to use NPV methods as rate of return methods such as IRR in their capital budgeting decisions. Absent compelling evidence to select one objective function over the other, I focus on NPV.

Benchmark 1: Let k_u^* and k_d^* be the optimal levels of upstream and downstream capital investment if the transfer price is independent of the capital levels. The following first-order conditions define k_u^* and k_d^* :

$$v'(k_u^*) = \frac{-1 + \sum_{n=1}^N D_n \delta_n t_u}{\sum_{n=1}^N D_n \gamma_n \tau_u}, \quad (2)$$

$$v'(k_d^*) = \frac{-1 + \sum_{n=1}^N D_n \delta_n t_d}{\sum_{n=1}^N D_n \gamma_n \tau_d}. \quad (3)$$

The investment is not tax-neutral; both k_u^* and k_d^* are decreasing in the respective tax rates.⁵ This is a variation of the well-known result in Sandmo (1974). In the next two sections I examine the effect of the specific transfer pricing regulations on the firm's investment decision.

III. COMPARABLE PROFIT METHOD

I first examine the comparable profit method (CPM). Since 1994, the IRS has allowed this method, which allocates income for tax purposes based on the *rate of return* of a comparable uncontrolled firm rather than on the price of a comparable uncontrolled transaction. Specifically, the CPM transfer price equates the rate of return on the division's assets to the rate of return on assets of a comparable uncontrolled firm (Treasury Reg. §1.482-5(a)). That is, one works backward from the comparable rate of return to arrive at the appropriate transfer price.

I assume that a comparable firm exists. That is, I assume that economic conditions permit the simultaneous existence of vertically integrated and independently organized firms.⁶ I do not model the independent firm, but rather take its rate of return as given. In particular, the rate of return is independent of the vertically controlled firm's input choices and the amount of *ex post* discretion. These might be related if, for example, there were strategic interactions.

Upstream Division

The Treasury Regulations define the accounting rate of return under the CPM as operating profits divided by the book value of assets. The pre-tax operating profit for the upstream division in period n is $\gamma_n(T_n - v_u(k_u)) - \delta_n k_u$. Let R_n be the accounting rate of return of the comparable firm in period n . Equating $[\gamma_n(T_n - v_u(k_u)) - \delta_n k_u]/(\lambda_n k_u)$ to R_n yields a transfer price in period n of:

$$T_n = \frac{\lambda_n}{\gamma_n} R_n k_u + v_u(k_u) + \frac{\delta_n}{\gamma_n} k_u. \quad (4)$$

⁵ The derivative of the first-order conditions with respect to t_j is:

$$\frac{-(\sum_{n=1}^N D_n \gamma_n)(1 - \sum_{n=1}^N D_n \delta_n)}{(\sum_{n=1}^N D_n \gamma_n \tau_j)^2} < 0.$$

⁶ Independently organized firms are more susceptible to the hold-up problem in relationship-specific investment, but vertically integrated entities may have higher agency costs because of the potential separation of ownership from control. If the respective costs offset each other, then both types of organization can exist.

Because the firm makes its investment decisions at time 0, it must predict, based on its knowledge of comparable firms in its industry, the comparable rates of return ($R_1, R_2, \dots, R_N$) in future periods and the implied transfer prices ($T_1, T_2, \dots, T_N$). Discounting the stream of transfer payments in Equation (4) yields:

$$\sum_{n=1}^N D_t \gamma_n T_n = \sum_{n=1}^N D_n \lambda_n R_n k_u + \sum_{n=1}^N D_n \gamma_n v_u(k_u) + \sum_{n=1}^N D_n \delta_n k_u.$$

The discounted sum of the revenue transfers is partially composed of the discounted sum of the accounting depreciation expenses; in essence, the CPM regulations monetize the depreciation expenses. Let $R^* = \sum_{n=1}^N D_n \lambda_n R_n$. The variable R^* aggregates the firm's predictions about the future comparable rates of return. As the following expression for the discounted stream of transfer payments indicates, one can interpret R^* as a "cumulative" accounting rate of return on original capital:

$$\sum_{n=1}^N D_n \gamma_n T_n = R^* k_u + \sum_{n=1}^N D_n \gamma_n v_u(k_u) + \sum_{n=1}^N D_n \delta_n k_u. \quad (5)$$

Because it chooses the level of upstream capital, the firm has *ex ante* control over the present value of the transfer payments to the upstream division, as described in the following lemma. It is convenient for the following lemmas and propositions to define $\Delta = 1 - \sum_{n=1}^N D_n \delta_n$.

Lemma 1: If the comparable profit method is applied to the upstream division, then the present value of the transfer payments to the upstream division is increasing in k_u at $k_u = k_u^*$ if $\tau_u R^* > \Delta$, fixed if $\tau_u R^* = \Delta$, and decreasing if $\tau_u R^* < \Delta$.

The result holds for all accounting depreciation policies (the depreciation factor δ_n 's, which imply the book value λ_n 's) and patterns of economic usage (the γ_n 's). An increase in k_u lowers the rate of return by increasing the depreciation expense and the capital base, and increases the rate of return by lowering variable costs. Which effect dominates at k_u^* depends on the relative sizes of the after-tax return to a unit of investment ($\tau_u R^*$) and a term relating to the present value of the depreciation write-off (Δ). If $\tau_u R^*$ is sufficiently high, setting $k_u > k_u^*$ increases the discounted present value of the transfer payments. Otherwise, setting $k_u > k_u^*$ decreases it. The results in the CPM hinge on the relation between $\tau_u R^*$ and Δ . The next lemma analyzes it for a special case.

Lemma 2: If the comparable rates of return are constant ($R_n = R$, for all n), then the present value of the transfer payments to the upstream division is increasing in k_u at $k_u = k_u^*$ if $\tau_u R > i$, fixed if $\tau_u R = i$, and decreasing if $\tau_u R < i$.

In this special case, $\tau_u R^* = \Delta$ is equivalent to $\tau_u R = i$, regardless of the depreciation policy. That is, if the after-tax comparable rate of return is equal to the discount rate, then deviations from k_u^* do not affect the amount of income shifted across divisions via the transfer payments. In practice, i , the discount rate the firm uses for capital budgeting

purposes, and $\tau_u R$, the after-tax rate of a comparable firm, are not necessarily equal. Which is greater, however, likely depends on firm-specific factors, limiting the generality of the CPM results. I argue informally that one would expect $\tau_u R$ to be less than i , implying that $\tau_u R^* < \Delta$, for two reasons. First, the accounting rate of return R is likely to understate the true rate of return because accounting depreciation is generally more accelerated than economic depreciation.⁷ Second, anecdotal evidence suggests that firms set artificially high cut-off rates for capital projects (Brealey and Myers 1996, 220). If R_a varies, then the difference between $\tau_u R^*$ and Δ is still closely related to the difference between the average after-tax comparable rate of return and the discount rate, but the relation is not exact. Furthermore, it depends on the depreciation policy.

Lemma 1 shows that, unlike the benchmark case, the transfer price in this scenario may depend on k_u . Substituting the expression for the present value of transfer payments (Equation [5]) into the benchmark objective function (Equation [1]) yields a net present value of:

$$-(k_u + k_d) + \sum_{n=1}^N D_n \gamma_n [\tau_d (p - v_d(k_d)) - \tau_u v_u(k_u)] + \sum_{n=1}^N D_n \delta_n (t_u k_u + t_d k_d) + (\tau_u - \tau_d) \left[R^* k_u + \sum_{n=1}^N D_n \gamma_n v_u(k_u) + \sum_{n=1}^N D_n \delta_n k_u \right]. \quad (6)$$

The benchmark investments k_u^* and k_d^* maximize the sum of the expressions in the first line of the new objective function. The term on the second line represents the distortion to the firm's incentives induced by application of the regulations. Its sign depends on the relative tax rates and on the condition in Lemma 1. The next proposition compares the CPM investments to the benchmark investments.

Proposition 1: The comparable profit method, applied to the upstream division:

- (i) Induces overinvestment in upstream capital relative to the benchmark case if $t_d > t_u$ and $\tau_u R^* > \Delta$, or $t_u > t_d$ and $\tau_u R^* < \Delta$.
- (ii) Induces underinvestment in upstream capital relative to the benchmark case if $t_u > t_d$ and $\tau_u R^* > \Delta$, or $t_d > t_u$ and $\tau_u R^* < \Delta$.
- (iii) Induces no distortion in upstream capital if $\tau_u R^* = \Delta$.
- (iv) Induces no distortion in downstream capital.

Because of the assumption that the downstream capital has no effect on the upstream variable cost, the calculation of the upstream division's return depends only on upstream capital. Therefore, the method induces no distortion in k_u .⁸ If the upstream division has the lower tax rate, then the firm has incentive to shift income upstream with a higher

⁷ Consider a firm that buys the same asset every period. In the steady state in which the firm holds assets in every stage of useful life, the aggregate depreciation charge is the same regardless of the speed of depreciation. However, the aggregate net book value is higher if the depreciation is slower.

⁸ If the variable cost depended on both investments, then the extension would be straightforward. Suppose $(\partial v_u(\cdot) / \partial k_d) < 0$. Then, decreasing k_d would increase the upstream variable cost and hence the transfer price. Therefore, the firm would underinvest (overinvest) in k_d if $t_u < t_d$ ($t_u > t_d$).

transfer price. Lemma 1 shows that the firm can accomplish this by increasing upstream capital if R^* is sufficiently high, decreasing it otherwise. That is, the firm shifts income *ex ante* by distorting the investment choices. Based on the earlier discussion, I argue that underinvestment is the more plausible outcome. If the downstream division has the lower tax rate, then the firm shifts income downstream with lower transfer payments.

An interesting implication of Proposition 1 is that the CPM induces no productive distortions, regardless of tax rate differences, if $\tau_u R^* = \Delta$, a result that will not hold under the CUP method. As Lemma 2 shows, this is equivalent to $\tau_u R = i$ in the special case of constant R_n . *Ex ante* shifting incentives arise from the firm's ability to affect the present value of transfer payments by distorting the investment from k_u^* . If the after-tax return, discount rate, and depreciation schedules are exactly calibrated, then deviations from k_u^* do not affect the aggregate payments. Hence, there are no distortions.

Thus far, I have implicitly assumed that regulators can enforce the "true" arm's length price. The investment distortions in Proposition 1 are purely the function of *ex ante* income shifting. I now relax this assumption and explore the implications of *ex post* shifting. Each set of capital investments at time 0 implies the exact transfer price in Equation (4) in period n . However, the enforcement of transfer pricing regulations is imperfect. In the comparable profit method, the need to identify comparable firms and to adjust the rate of return to reflect material differences between the controlled and uncontrolled entities are significant practical challenges (Treasury Reg. §1.482-1[d](2)). In practice, tax auditors calculate an arm's length range based on multiple comparables, making no transfer pricing adjustment if the firm's operating profits fall within this range.⁹

I implement this aspect of the institutional environment by assuming that the firm can defend a rate of return in period n of $R_n + \phi$, with $\phi \in [-\Phi, \Phi]$.¹⁰ The variable Φ represents the maximum *ex post* discretion the firm has over the transfer price. The introduction of *ex post* discretion implies that the transfer price in period n is:

$$\left[\frac{\lambda_n}{\gamma_n} R_n k_u + v_u(k_u) + \frac{\delta_n}{\gamma_n} k_u \right] + \phi \frac{\lambda_n}{\gamma_n} k_u. \quad (7)$$

As a result, *ex post* discretion maps into the following range of prices in period n :

$$v_u(k_u) + k_u \left[\frac{\lambda_n}{\gamma_n} (R_n - \Phi) + \frac{\delta_n}{\gamma_n} \right] \leq T_n \leq v_u(k_u) + k_u \left[\frac{\lambda_n}{\gamma_n} (R_n + \Phi) + \frac{\delta_n}{\gamma_n} \right]. \quad (8)$$

If $\Phi = 0$, then the range collapses to the expression in Equation (4). I assume that the firm knows with certainty the amount of *ex post* discretion tolerable (or undetectable) to tax regulators. Implicitly, the penalties associated with violating this bound are severe enough that the firm never exceeds it. The modeling captures both dimensions of discretion. First, a firm, via its input choices, determines the standard of comparability. Second, even with

⁹ Treasury Reg. §1.482-1[e][2][iii](A). More precisely, the entire range is acceptable if the comparable data are reliable. If not, the acceptable range is the interquartile range. (Treasury Reg. §1.482-1[e][2][iii](B)).

¹⁰ Clearly, Φ must be small relative to R_n . I also assume that the maximum discretion is not so large that the firm can choose a transfer price producing losses in either division.

a given standard of comparability, the auditors' estimate of the arm's length price is imprecise. Although the certainty assumption simplifies the model, the results do not depend on the firm's knowing at time 0 the level of *ex post* discretion in future periods.¹¹

The firm's income shifting is a sequential process. In periods 1 through n the firm, given its choice of k_u and k_d at time 0, chooses the *ex post* transfer price (i.e., chooses the ϕ) that minimizes the taxes of:

$$t_d(p - T_n - v_d(k_d)) + t_u(T_n - v_u(k_u)). \quad (9)$$

Substitution of the expression for T_n in Equation (7) into Equation (9) and differentiation with respect to ϕ yields $(\lambda_n/\gamma_n) k_u(t_u - t_d)$. If $t_d > t_u$, then the tax is decreasing in ϕ and the firm sets the highest transfer price: $\phi = \Phi$. The reverse holds if $t_u > t_d$.

By the definition of comparability, the discounted sum of the portion of the transfer price in brackets in Equation (7) is the expression in Equation (5). The discounted sum of the portion of each period's transfer price attributable to *ex post* discretion is $\sum_{n=1}^N D_n \lambda_n \phi k_u$. Let $\mu = \sum_{n=1}^N D_n \lambda_n$. The firm's ability to deviate *ex post* from the true arm's length price generates extra aggregate return on capital of $\mu\phi$. The last term of the objective function (Equation [6]) is now:

$$(\tau_u - \tau_d) \left[(R^* \pm \mu\Phi)k_u + \sum_{n=1}^N D_n \gamma_n v_u(k_u) + \sum_{n=1}^N D_n \delta_n k_u \right],$$

with the sign on $\mu\Phi$ depending on the relative tax rates. Anticipating the optimal selection of ϕ in subsequent periods, the firm selects at time 0 the k_u and k_d that maximize the net present value of the cash flows. The next proposition shows the effect of *ex post* discretion on the distortions to investment.

Proposition 2: If the comparable profit method, applied to the upstream division, induces overinvestment, then increasing *ex post* discretion (Φ) exacerbates the distortion. If it induces underinvestment, then increasing the *ex post* discretion mitigates the distortion.

Note that the introduction of *ex post* discretion alters the conditions in Proposition 1 because it affects the after-tax return on an incremental unit of capital. If $t_d > t_u$, then the

¹¹ Suppose that instead of knowing with certainty the limit of *ex post* discretion, the firm knows only the expected penalty associated with deviating from the true arm's length price. Let this penalty function be $P(\phi, \eta)$ with $P_\phi, P_{\phi\phi}, P_{\phi\eta}$ all positive. The parameter η captures the severity of the regulatory regime. Increasing η shifts P_ϕ higher for all ϕ , corresponding to a regime with lower *ex post* discretion. The reduced-form modeling is consistent with Kant (1988). In this case the firm would minimize the sum of taxes and expected penalties of:

$$t_d(p - T_n - v_d(k_d)) + t_u(T_n - v_u(k_u)) + P(\phi, \eta)$$

and the optimal amount of discretion would be determined implicitly by the first-order condition:

$$k_u(t_d - t_u) = P_\phi.$$

Increasing η is equivalent to decreasing Φ in the model. As P_ϕ increases, the firm sets an *ex post* transfer price closer to the true price. The optimal ϕ would feed back into the firm's time 0 decision in the same manner. The richer model of *ex post* discretion, therefore, does not lead to different insights into the basic interaction between *ex ante* and *ex post* discretion, the primary focus of this paper.

condition on the comparable rate of return is $\tau_u(R^* + \mu\Phi) > \Delta$; if $t_u > t_d$, then it is $\tau_u(R^* - \mu\Phi) > \Delta$. To understand the proposition, consider a scenario in which the upstream division has the lower tax rate and $\tau_u(R^* + \mu\Phi) > \Delta$. Because $t_d > t_u$, shifting income via a high transfer price is desirable. Lemma 1 implies that selecting $k_u > k_u^*$ shifts the future defensible range of transfer prices higher. Furthermore, *ex post* discretion reinforces *ex ante* incentives, exacerbating the production distortion. In this case, the transfer price is equal to the upper bound of the range in Equation (8). The amount by which increases in upstream investment raise the transfer price is increasing in *ex post* discretion. Hence, the firm distorts k_u even further if Φ increases. Conversely, if $t_u > t_d$ and $\tau_u(R^* - \mu\Phi) < \Delta$, shifting income downstream via a low transfer price is optimal. Selecting $k_u > k_u^*$ shifts the future defensible range of transfer prices lower. The firm chooses the lower bound of the range as the transfer price. The amount by which decreases in upstream investment lower this transfer price is decreasing in discretion. Hence, the ability to shift income *ex post* mitigates the *ex ante* incentives. The firm reduces the distortion in k_u as *ex post* discretion increases.

Previous transfer pricing papers demonstrate that arm's length regulations distort the firm's decisions. Proposition 2 shows that if one considers both forms of income shifting simultaneously, then *ex post* shifting, perceived to be a practice that serves only to undermine regulations, may have the benefit of alleviating production distortions. In fact, the proposition suggests that the tax revenues of both jurisdictions increase in the amount of *ex post* discretion if the more efficient input selections increase the size of pre-tax profits enough to offset the higher tax-rate division's lower share of those profits.

The firm deviates within, rather than from, the transfer pricing rule. If discretion implied that the firm could choose an arbitrary transfer price, then the regulation-induced investment distortion would vanish in the limit. As discretion increased, the firm's objective function would eventually converge to the benchmark case (Equation [1]) in which the transfer price is independent of the capital levels. The required adherence to the method of calculating the transfer price produces the asymmetry in the comparative statics.

Downstream Division

The analysis proceeds in a symmetric way if one applies the comparable price method to the downstream jurisdiction. Pre-tax operating profits for the downstream division in period n are $\gamma_n(p - v_d(k_d) - T_n) - \delta_n k_d$. Equating the rate of return to the predicted R_n 's and aggregating yields the following expression for the present value of the transfer payments to the upstream division as a function of downstream investment k_d :

$$\sum_{n=1}^N D_n \gamma_n T_n = \sum_{n=1}^N D_n \gamma_n (p - v_d(k_d)) - \sum_{n=1}^N D_n \delta_n k_d - R^* k_d, \quad (10)$$

where R^* is defined as in the previous section. In this scenario, the firm controls the present value of the transfer payments to the upstream division by changing downstream investment.

Lemma 3: If the comparable profit method is applied to the downstream division, then the present value of the transfer payments to the upstream division is decreasing in k_d at $k_d = k_d^*$ if $\tau_d R^* > \Delta$, fixed if $\tau_d R^* = \Delta$, and increasing if $\tau_d R^* < \Delta$. In the special case of constant $R_n = R$, the discounted transfer payments are decreasing in k_d at $k_d = k_d^*$ if $\tau_d R > i$, fixed if $\tau_d R = i$, and increasing if $\tau_d R < i$.

This lemma combines the results of Lemmas 1 and 2 for the upstream division. The sign on the condition reverses because a higher downstream return corresponds to a lower transfer price. Substituting the expression for the present value of transfer payments (Equation [10]) into the benchmark objective function (Equation [1]) yields a net present value of:

$$-(k_u + k_d) + \sum_{n=1}^N D_n \gamma_n [\tau_d(p - v_d(k_d)) - \tau_u v_u(k_u)] + \sum_{n=1}^N D_n \delta_n (t_u k_u + t_d k_d) + (\tau_u - \tau_d) \left[\sum_{n=1}^N D_n \gamma_n (p - v_d(k_d)) - \sum_{n=1}^N D_n \delta_n k_d - R^* k_d \right]. \quad (11)$$

The benchmark investments k_u^* and k_d^* maximize the sum of the expressions in the first line of the new objective function. The last term represents the distortion to the firm's incentives induced by application of the regulations. Its sign depends on the relative tax rates and on the condition in Lemma 3. The next proposition compares the CPM investments to the benchmark investments in the pure *ex ante* setting.

Proposition 3: The comparable profit method, applied to the downstream division:

- (i) Induces underinvestment in downstream capital relative to the benchmark case if $t_d > t_u$ and $\tau_d R^* > \Delta$, or $t_u > t_d$ and $\tau_d R^* < \Delta$.
- (ii) Induces overinvestment in downstream capital relative to the benchmark case if $t_u > t_d$ and $\tau_d R^* > \Delta$, or $t_d > t_u$ and $\tau_d R^* < \Delta$.
- (iii) Induces no distortion in downstream capital if $\tau_d R^* = \Delta$.
- (iv) Induces no distortion in upstream capital.

As in the upstream case, tax rate differentials need not distort investments if $\tau_d R^*$ and Δ calibrate exactly. As I discussed in the previous section, $\tau_d R^* < \Delta$ is the most plausible scenario, but the opposite could hold for firm-specific reasons. The modeling of *ex post* discretion is also similar. In period n , the CPM transfer price is:

$$T_n = p - v_d(k_d) - \frac{\delta_n}{\gamma_n} k_d - \frac{\lambda_n}{\gamma_n} k_d R_n. \quad (12)$$

The introduction of *ex post* discretion implies that the transfer price in period n is:

$$\left[p - v_d(k_d) - \frac{\delta_n}{\gamma_n} k_d - \frac{\lambda_n}{\gamma_n} k_d R_n \right] - \phi \frac{\lambda_n}{\gamma_n} k_d. \quad (13)$$

Maximum discretion of Φ implies a range of prices of:

$$p - v_d(k_d) - \frac{\delta_n}{\gamma_n} k_d - \frac{\lambda_n}{\gamma_n} k_d (R_d + \Phi) \leq T_n \leq p - v_d(k_d) - \frac{\delta_n}{\gamma_n} k_d - \frac{\lambda_n}{\gamma_n} k_d (R_d - \Phi). \quad (14)$$

Higher *ex post* discretion implies that the firm can choose from a wider range of defensible

prices. Also, the firm's time 0 capital investment choices shift the interval higher or lower. By the definition of comparability, the discounted sum of the portion of the transfer price in brackets in Equation (13) is the expression in Equation (10). The last term of the objective function is now:

$$(\tau_u - \tau_d) \left[\sum_{n=1}^N D_n \gamma_n (p - v_d(k_d)) - \sum_{n=1}^N D_n \delta_n k_d - (R^* \pm \mu \Phi) k_d \right],$$

with μ defined as before and the sign on $\mu\Phi$ depending on the relative tax rates. The next proposition shows the effect of *ex post* discretion on the distortions to investment.

Proposition 4: If the comparable profit method, applied to the upstream division, induces overinvestment, then increasing *ex post* discretion (Φ) exacerbates the distortion. If it induces underinvestment, then increasing the *ex post* discretion mitigates the distortion.

Again, the introduction of *ex post* discretion alters the conditions in Proposition 3. If $t_d > t_u$, the condition on the comparable rate of return is $\tau_d(R^* - \mu\Phi) > \Delta$; otherwise the condition is $\tau_d(R^* + \mu\Phi) > \Delta$. Proposition 4 is symmetric to Proposition 2. *Ex post* discretion may relieve *ex ante* distortions, again suggesting the possibility of higher tax revenues in both jurisdictions resulting from higher *ex post* discretion.

IV. CUP METHOD

I next examine the comparable uncontrolled price (CUP) method of determining an arm's length transfer price. Unlike the CPM, which infers the appropriate transfer price from the rate of return on assets of unrelated parties, the CUP method derives the transfer price from similar or identical transactions involving unrelated parties (Treasury Reg. §1.482-3(b)). That is, the divisions must transact at the same price that obtains in a similar transaction between independent firms. Prior to 1994, regulations gave priority to the CUP method over other methods. The 1994 regulations diminished its dominance.

I implement the CUP method analogously to Sansing (1999). Suppose that, after making the time 0 investments, the firm sold one of the divisions to an uncontrolled party. The CUP method asks, "At what price would these (now) unrelated parties transact?" Because of the relationship-specific assets, trading with each other is more beneficial than trading with external parties because each division enjoys a lower variable cost. The transaction price would depend on the surplus created by internal trade. Even though it is optimal to trade internally, the price would also depend on the external opportunities for trade, which define the negotiation "threat" points for the divisions.

The investments k_u and k_d are relationship-specific and have value only if the divisions trade internally. Because the initial investments k_u and k_d are sunk costs after time 0, the depreciation expenses $\delta_n k_u$ and $\delta_n k_d$ are not relevant for the calculation of the CUP price. The variable costs of production are $k_u(0) = \bar{k}_u$ and $k_d(0) = \bar{k}_d$, respectively, if the divisions trade externally. The upstream division can sell the intermediate good outside the firm at a price of x_u /unit. The upstream division, then, would realize pre-tax profits of $x_u - \bar{k}_u$ if it traded externally. The downstream firm can buy the intermediate good externally for x_d /unit, where $x_d \geq x_u$, and would realize pre-tax profits of $p - x_d - \bar{k}_d$ if it traded externally. Total pre-tax profits if both divisions traded externally, then, would be:

$$p + x_u - x_d - \bar{k}_d - \bar{k}_u. \quad (15)$$

Subtracting pre-tax profits from external trade (Equation [15]) from pre-tax profits under internal trade ($p - v_u(k_u) - v_d(k_d)$) yields the surplus generated by internal trade:

$$\bar{k}_u + \bar{k}_d - v_d(k_d) - v_u(k_u) - x_u + x_d > 0.$$

I assume that the divisions would split the surplus equally in the hypothetical negotiation. Hence, the transfer price T must satisfy:

$$T - v_u(k_u) - (x_u - \bar{k}_u) = \frac{1}{2} [\bar{k}_u + \bar{k}_d - v_d(k_d) - v_u(k_u) - x_u + x_d]$$

and

$$p - T - v_d(k_d) - (p - x_d - \bar{k}_d) = \frac{1}{2} [\bar{k}_u + \bar{k}_d - v_d(k_d) - v_u(k_u) - x_u + x_d],$$

implying that:

$$T = \frac{1}{2} [X - v_d(k_d) + v_u(k_u)], \quad (16)$$

where $X = \bar{k}_d - \bar{k}_u + x_u + x_d$.

I assume that x_d and x_u , and therefore X , are constant across time.¹² Unlike the CPM, in which the transfer price depends on the accounting depreciation expense and asset book value each period, the CUP transfer price is constant over time.

The CUP transfer price is decreasing in upstream capital and increasing in downstream capital. An increase in k_u , holding the transfer price constant, increases the surplus from internal trade by the amount of the decrease in upstream variable cost. To allocate half of the incremental surplus to the downstream division, the transfer price must decrease by half of the amount of the variable cost reduction. A symmetric argument holds for the effect of changes in the downstream division's capital investment on the transfer price.

The present value of the transfer payment is:

$$\sum_{n=1}^N D_n \gamma_n T = \frac{1}{2} \sum_{n=1}^N D_n \gamma_n [X - v_d(k_d) + v_u(k_u)].$$

Substituting this expression into the benchmark objective function (Equation [1]) yields a net present value of:

$$\begin{aligned} & -(k_u + k_d) + \sum_{n=1}^N D_n \gamma_n [\tau_d(p - v_d(k_d)) - \tau_u v_u(k_u)] + \sum_{n=1}^N D_n \delta_n (t_u k_u + t_d k_d) \\ & + \frac{1}{2} (\tau_u - \tau_d) \sum_{n=1}^N D_n \gamma_n [X - v_d(k_d) + v_u(k_u)]. \end{aligned} \quad (17)$$

The benchmark investments k_u^* and k_d^* maximize the sum of the terms of the first line of

¹² Because x_d and x_u are independent of k_u and k_d , the results would hold even if they were stochastic.

the objective function. The term on the second line represents the distortion to the firm's incentives induced by application of the regulations. Its sign depends on the relative tax rates. The next proposition describes the underinvestment and overinvestment CUP method induces.

Proposition 5: The comparable uncontrolled price method:

- (i) Induces underinvestment in upstream capital and overinvestment in downstream capital relative to the benchmark if $t_d > t_u$.
- (ii) Induces overinvestment in upstream capital and underinvestment in downstream capital relative to the benchmark if $t_u > t_d$.

If $t_d > t_u$, then the firm has incentive to shift income upstream with high future transfer prices. The firm increases the transfer price by overinvesting in downstream capital and underinvesting in upstream capital. The direction of the investment distortions reverses if $t_u > t_d$. Unlike with the CPM, changes in k_u or k_d do not have conflicting effects on the transfer price. Therefore, the under- and overinvestment in Proposition 5 depends only on the relative tax rates, unlike in Propositions 1 and 3.

In practice, regulations allow firms to group transactions when applying the CUP method (Treasury Reg. §1.482-1[f][2](iv)), which, as in the CPM, produces a range of acceptable transfer prices. Hence, I model *ex post* discretion over the comparable uncontrolled price transfer price with the assumption that $T = \frac{1}{2}(1 + \phi)[X - v_d(k_d) + v_u(k_u)]$, where $\phi \in [-\Phi, \Phi]$. This translates into a range of defensible transfer prices of:

$$\frac{1}{2}(1 - \Phi)[X - v_d(k_d) + v_u(k_u)] \leq T_n \leq \frac{1}{2}(1 + \Phi)[X - v_d(k_d) + v_u(k_u)]. \quad (18)$$

An increase in *ex post* shifting expands the range. The firm can shift the interval higher (lower) *ex ante* by increasing k_d (k_u).

Substitution of T into the expression for taxes in Equation (9) and differentiation with respect to ϕ yields $\frac{1}{2}(t_u - t_d)(X - v_d(k_d) + v_u(k_u))$. Tax minimization requires that the firm set the highest transfer price ($\phi = \Phi$) if $t_d > t_u$ and the lowest transfer price ($\phi = -\Phi$) otherwise. The last term in the objective function is now:

$$\frac{1}{2}(1 \pm \Phi)(\tau_u - \tau_d) \sum_{n=1}^N D_n \gamma_n [X - v_d(k_d) + v_u(k_u)],$$

where Φ depends on the relative tax rates. The next proposition discusses the effect of *ex post* discretion on investment decisions under the CUP method.

Proposition 6: Under the comparable uncontrolled price method, the distortion in both investments is increasing in the amount of *ex post* discretion Φ if $t_d > t_u$ and decreasing if $t_u > t_d$.

If $t_d > t_u$, then the transfer price is the upper end of the range in Equation (18). Clearly, the amount by which distortions in investments increase in transfer price is increasing in *ex post* discretion. At the other of the range (if $t_u > t_d$), increasing Φ reduces the ability to

shift the transfer price by changing the investments. Therefore, as with the CPM, it is possible that increases in the firm's *ex post* discretion could increase the tax revenues of both jurisdictions if $t_u > t_d$.

V. DISCUSSION

Samuelson (1982) is the first to model distortionary *ex ante* income shifting, but his model of transfer pricing is inconsistent with existing regulations. Specifically, the firm need not transfer at the existing comparable uncontrolled price. Subsequent research, such as Halperin and Srinidhi (1987, 1996) and Harris and Sansing (1998), model both the CUP method and the CPM specified by §482 Treasury Regulations. The settings in these papers differ, but they all yield the same basic result: Firms can distort the inputs from the no-tax optimal levels to obtain a more favorable regulatory transfer price. Propositions 1, 3, and 5 essentially replicate this earlier work, with the incremental contribution that the direction of the distortion under the CPM depends not only on the relative tax rates, but also on the relative magnitudes of the expected cumulative accounting rate of return and the present value of the depreciation tax benefit. This dependence arises because of the link between the CPM transfer price and the firm's accounting rate of return (and therefore the accounting depreciation policy). In the CUP analyses, the direction of the distortion depends only on the relative tax rates.

The Halperin and Srinidhi (1987, 1996) and Harris and Sansing (1998) papers suppress *ex post* discretion, implicitly assuming that the tax authority can enforce the regulatory transfer price. The Samuelson (1982) model does incorporate *ex post* shifting by allowing the firm to choose any price between its marginal cost of production and the price at which it sells the intermediate product domestically. The firm optimally chooses an extreme transfer price, as in my paper. Inferences from Samuelson's (1982) model are limited, however, by the fact that his model is inconsistent with current regulations. Moreover, his study does not consider the simultaneous interaction of *ex ante* and *ex post* shifting. That is, he does not analyze the effect of varying the width of the *ex post* range on the *ex ante* input distortions.

Kant (1988), on the other hand, focuses exclusively on *ex post* shifting. The firm can deviate from the arm's length price, but at some risk of detection and punishment by tax authorities. The firm does not always choose the most extreme transfer price, a result absent in my model.¹³ By varying the probability of detection for a given deviation from the arm's length price, Kant (1988) varies *ex post* discretion. Changes in *ex post* discretion affect the firm's input choices. The Kant model does not allow *ex ante* income shifting, however, because the arm's length price is exogenous (independent of the firm's inputs). Because the firm cannot affect the arm's length transfer price, it produces efficiently even in the presence of tax rate differentials if *ex post* shifting is not available. This is in contrast to my Propositions 1, 3, and 5, which show that tax rate differentials induce input distortions under current regulations even in the absence of *ex post* discretion.

Propositions 2, 4, and 6, which simultaneously consider the effects of *ex ante* and *ex post* discretion in an institutionally valid context, represent the principal contribution of my study. The main insight from considering the feedback between the two is that permitting firms *ex post* discretion in selecting transfer prices may result in more efficient investment. This result is potentially important to the frequent debates about transfer pricing policy. Given my focus on the efficiency implications of transfer pricing policy, I discuss the

¹³ As I discuss in footnote 11, I could obtain this result by modeling enforcement less starkly.

welfare effects only informally. In the most plausible setting, increasing *ex post* discretion would have an ambiguous effect on welfare, but there is at least the theoretical possibility of unambiguous effects. This possibility exists because *ex post* shifting may relieve *ex ante* inefficiencies.¹⁴

VI. CONCLUSION

This paper examines the interaction between two types of income shifting in multinationals. Transfer pricing regulations link the firm's allocation of income across jurisdictions to the firm's input decisions. By distorting the inputs, the firm can generate a more favorable "true" transfer price for taxation purposes. I label this *ex ante* income shifting. Imperfect enforcement of transfer pricing regulations means the firm can choose a transfer price that deviates from the true arm's length price. I label this *ex post* income shifting. In other words, *ex post* discretion implies a range of future defensible transfer prices; *ex ante* inputs determine the central tendency of the range. Previous research considers either one of the two modes of income shifting in isolation, but not the interaction between the two.

I examine both the comparable profit method (CPM) and the comparable uncontrolled price (CUP) method of calculating an arm's length transfer price. The study's main result is that *ex post* shifting may mitigate *ex ante* investment distortions. If the firm can deviate more from the true transfer price (higher *ex post* discretion), then it may distort the original inputs less. *Ex post* income shifting replaces *ex ante* shifting, and the reduction in *ex ante* shifting has efficiency implications. Depending on the transfer pricing method used and the relative tax rates, *ex post* shifting either exacerbates or alleviates *ex ante* investment distortions. A complete analysis of *ex post* income shifting must consider its effect on the *ex ante* input distortions. Although the improved efficiency can occur under either the CUP method or the CPM, the CUP method yields the sharpest results. In the CPM, the direction of the investment distortion depends on the discount rate and depreciation policy, both of which are idiosyncratic to the firm. In the CUP method, the direction of distortion depends only on tax-rate differentials.

Although I focus on the efficiency effects of the regulations, the results suggest some distributional implications. In particular, Propositions 2, 4, and 6 imply that tax revenues in both jurisdictions could increase in the level of *ex post* discretion if the increased pre-tax profits resulting from more efficient investment decisions more than offset the higher tax-rate jurisdiction's lower share of those profits. That is, a smaller piece of a bigger pie might lead to higher tax collections. A full analysis of the distributional effects of *ex post* discretion would require modeling of the objective function and actions of tax authorities, which is beyond the scope of this study. Efficiency is not the only goal of regulators. While higher *ex post* discretion may improve efficiency, the improved efficiency may be less important to regulators than maintaining the income redistribution that the transfer pricing regulations were presumably intended to create. Only a full equilibrium analysis can derive the optimal amount of *ex post* discretion.

The study is also a partial equilibrium analysis in that instead of explicitly modeling the economic conditions under which both independent and vertically integrated firms exist, I impose the coexistence exogenously. This coexistence is important. If vertically integrated firms dominate independent firms, then there are no uncontrolled transactions from which to infer arm's length prices. If independent firms dominate vertically integrated firms, then

¹⁴ The absence of feedback between *ex ante* and *ex post* shifting explains the ambiguous welfare results in Kant (1988). In his paper, the improvement in welfare of one jurisdiction always comes at the expense of the other.

transfer prices do not exist. I avoid explicitly modeling coexistence by taking as given the comparable rate of return in the CPM and the external opportunities for trade in the CUP method, and by suppressing agency considerations.

Although only a partial equilibrium analysis, this study helps to rationalize the perceived lax enforcement, or unenforceability, of transfer pricing regulations. When determining the specificity of regulations and the level of resources to allocate to enforcement, tax authorities likely trade off tax revenues, productive efficiency, and auditing and political costs. Although *ex post* shifting may conflict with the other goals, it may improve *ex ante* productive efficiency. This result could explain a relatively lax regulatory regime.

APPENDIX A NOTATION

TABLE 1
Notation

<i>Term</i>	<i>Definition</i>
p	price of the finished good
k_u	upstream capital investment
k_d	downstream capital investment
q	quantity produced
N	economic life of asset
$v_j(k_j)$	variable cost function for division j
γ_n	economic depreciation factor in period n
δ_n	accounting depreciation factor in period n
λ_n	book value of asset at beginning of period n as a percentage of historical cost
T_n	transfer price in period n
t_u	selling division (upstream) tax rate
t_d	buying division (downstream) tax rate
τ_j	$1 - t_j$, where $j = u, d$ (upstream or downstream division)
i	the firm's discount rate
D_n	discount factor in period n
R_n	accounting rate of return for comparable firm in period n
R^*	cumulative accounting rate of return ($\sum_{n=1}^N D_n \lambda_n R_n$)
Δ	factor related to depreciation expense ($1 - \sum_{n=1}^N D_n \delta_n$)
ϕ	<i>ex post</i> discretion in setting transfer prices
Φ	maximum absolute value of <i>ex post</i> discretion
μ	return multiplier on <i>ex post</i> discretion (CPM)
x_d	external purchase price for downstream division (CUP)
x_u	external sale price for upstream division (CUP)

APPENDIX B PROOFS

Proof of Benchmark Case

Let NPV_{uu} be the second derivative of the objective function with respect to k_u and let NPV_{dd} be the second derivative of the objective function with respect to k_d . Because the transfer price is independent of k_u and k_d , $NPV_{uu} = -\tau_u v''_u(k_u) \sum_{n=1}^N D_n \gamma_n < 0$ and $NPV_{dd} = -\tau_d v''_d(k_d) \sum_{n=1}^N D_n \gamma_n < 0$. $NPV_{ud} = NPV_{du} = 0$. Therefore, the objective function is concave and the first-order conditions characterize the maximizing values of k_u and k_d . The expressions in the paper are implied by the first-order conditions:

$$\begin{aligned} -1 - \sum_{n=1}^N D_n \gamma_n v'_u(k_u) \tau_u + \sum_{n=1}^N D_n \delta_n t_u &\doteq 0 \\ -1 - \sum_{n=1}^N D_n \gamma_n v'_d(k_d) \tau_d + \sum_{n=1}^N D_n \delta_n t_d &\doteq 0. \end{aligned}$$

Proof of Lemma 1

The derivative of the present value of the transfer payments with respect to k_u evaluated at $k_u = k_u^*$ (the substitution in line 2) is:

$$\begin{aligned} \frac{\partial}{\partial k_u} \left[R^* k_u + \sum_{n=1}^N D_n [\gamma_n v_u(k_u) + \delta_n k_u] \right] &= R^* + \sum_{n=1}^N D_n \delta_n + \sum_{n=1}^N D_n \gamma_n v'_u(k_u) \\ &= R^* + \sum_{n=1}^N D_n \delta_n + \frac{1}{\tau_u} \left[-1 + \sum_{n=1}^N D_n \delta_n t_u \right] \\ &= R^* - \frac{1}{\tau_u} \left[1 - \sum_{n=1}^N D_n \delta_n \right]. \end{aligned}$$

Therefore the derivative is positive (negative) if $\tau_u R^* > \Delta$ ($\tau_u R^* < \Delta$), where $\Delta = 1 - \sum_{n=1}^N D_n \delta_n$. ■

Proof of Lemma 2

I will prove the claim that $\tau_u R^* = \Delta$ if $\tau_u R = i$ by mathematical induction. The following table recalls the definitions of the variables in the paper.

n	D_n	δ_n	λ_n	$D_n \lambda_n R$	$D_n \delta_n$
1	$(1+i)^{-1}$	δ_1	1	$R(1+i)^{-1}$	$(1+i)^{-1} \delta_1$
2	$(1+i)^{-2}$	δ_2	$1 - \delta_1$	$R(1+i)^{-2}(1 - \delta_1)$	$(1+i)^{-2} \delta_2$
3	$(1+i)^{-3}$	δ_3	$1 - \delta_1 - \delta_2$	$R(1+i)^{-3}(1 - \delta_1 - \delta_2)$	$(1+i)^{-3} \delta_3$
n	$(1+i)^{-n}$	δ_n	δ_n	$R(1+i)^{-n} \delta_n$	$(1+i)^{-n} \delta_n$

Because the D_n 's must sum to 1, $\delta_n = 1 - \delta_1 - \delta_2 - \dots - \delta_{n-1}$. I will use this substitution to obtain Equations (B.5) and (B.6) below. For arbitrary n , then, $R^* = \sum_{n=1}^N D_n \lambda_n R$ is $[R/(1+i)] + [R(1-\delta_1)/(1+i)^2] + \dots + [R(1-\delta_1-\delta_2-\dots-\delta_{n-2})/(1+i)^{n-1}] + [(R\delta_n)/(1+i)^n]$, which is equivalent to:

$$R \frac{(1+i)^{n-1} + (1+i)^{n-2}(1-\delta_1) + \cdots + (1+i)(1-\delta_1-\delta_2-\cdots-\delta_{n-2}) + \delta_n}{(1+i)^n} \quad (\text{B.1})$$

For arbitrary n , $\sum_{n=1}^N D_n \delta_n$ is $[\delta_1/(1+i)] + [\delta_2/(1+i)^2] + \cdots + [\delta_n/(1+i)^n]$ and $(1/\tau_u) \Delta = (1/\tau_u) (1 - \sum_{n=1}^N D_n \delta_n)$ is:

$$\frac{1}{\tau_u} \frac{(1+i)^n - (1+i)^{n-1}\delta_1 - (1+i)^{n-2}\delta_2 - \cdots - (1+i)\delta_{n-1} - \delta_n}{(1+i)^n} \quad (\text{B.2})$$

I want to show that Equation (B.1) equals (B.2) if $R = i/\tau_u$. After substituting $R = i/\tau_u$ and factoring out $1/((1+i)\tau_u)$, one obtains the following expressions for Equations (B.1) and (B.2):

$$i[(1+i)^{n-1} + (1+i)^{n-2}(1-\delta_1) + \cdots + (1+i)(1-\delta_1-\delta_2-\cdots-\delta_{n-2}) + \delta_n], \quad (\text{B.3})$$

and

$$(1+i)^n - (1+i)^{n-1}\delta_1 - \cdots - (1+i)\delta_{n-1} - \delta_n. \quad (\text{B.4})$$

As the basis for the mathematical induction, I assume that the difference between Equations (B.3) and (B.4) is 0 for the n case. The difference is a first-order polynomial in $\delta_1, \delta_2, \dots, \delta_{n-1}$. For the difference to be 0 for any possible set of δ_n 's the coefficient on each δ_n and on the intercept must be 0. After making the proper substitutions for δ_n , the coefficients on the intercept and δ_1 , respectively, are:

$$i(1+i)^{n-1} + i(1+i)^{n-2} + \cdots + i(1+i) + i - [(1+i)^n - 1] = 0 \quad (\text{B.5})$$

$$-i(1+i)^{n-2} - i(1+i)^{n-3} - \cdots - i(1+i) - i + [(1+i)^{n-1} - 1] = 0 \quad (\text{B.6})$$

Now consider the case of $n+1$. The coefficients on δ_1 and δ_2 are, respectively:

$$-i(1+i)^{n-1} - i(1+i)^{n-2} - \cdots - i(1+i) - i + [(1+i)^n - 1] = 0 \quad (\text{B.7})$$

$$-i(1+i)^{n-2} - i(1+i)^{n-3} - \cdots - i(1+i) - i + [(1+i)^{n-1} - 1] = 0 \quad (\text{B.8})$$

Note that Equation (B.8) is equivalent to Equation (B.6), which by assumption is equal to 0. Similarly, the coefficients on δ_3 through δ_n in the $n+1$ case are equivalent to the coefficients on δ_2 through δ_{n-1} in the n case, and are therefore also 0. Therefore, I need only show that the coefficients on δ_1 and the intercept are 0 for the $n+1$ case. First, I show this for δ_1 . Rearrange Equation (B.6), the coefficient on δ_1 in the n case as:

$$(1+i)[(1+i)^{n-2} - i(1+i)^{n-3} - i(1+i)^{n-4} - \cdots - i] - 1 - i.$$

By the induction assumption, this equals 0. Therefore, the term in square brackets must be 1. Now rearrange the coefficient on δ_1 for the $n+1$ case (Equation (B.7)) as:

$$(1 + i)\{(1 + i)[(1 + i)^{n-2} - i(1 + i)^{n-3} - i(1 + i)^{n-4} - \dots - i] - i\} - 1 - i.$$

The previous argument establishes that the term in square brackets is 1. Therefore, the term in braces is also 1, and the entire term is 0. I omit the proof, which is similar, that the coefficient on the intercept is also 0 in the $n + 1$ case. Therefore, if the claim holds for n , it also holds for $n + 1$.

Now, for $n = 1$. For $n = 1$, the derivative is equal to $(1/(1 + i))R - (1/\tau_u)[1 - 1/(1 + i)]$. This clearly equals 0 if $R = i/\tau_u$. Therefore, by mathematical induction, the claim is true. ■

Proof of Proposition 1

Let the objective in Equation (6) be Π . $\Pi_{uu} = -\sum_{n=1}^N D_n \gamma_n \tau_d v_u''(k_u) < 0$, $\Pi_{dd} = -\sum_{n=1}^N D_n \gamma_n \tau_u v_d''(k_d) < 0$, and $\Pi_{ud} = \Pi_{du} = 0$. Therefore, the function is concave. The first derivative with respect to k_u is:

$$\left[-1 - \sum_{n=1}^N D_n \gamma_n v_u'(k_u) \tau_u + \sum_{n=1}^N D_n \delta_n t_u \right] + (t_d - t_u) \left[R^* + \sum_{n=1}^N D_n [\gamma_n v_u'(k_u) + \delta_n] \right].$$

At k_u^* , the first term in large brackets is 0. The sign of the derivative therefore depends on both the relative tax rates and the sign of the second term in large brackets (the distortion term). As Lemma 1 establishes, the last term in large brackets is positive at k_u^* if $\tau_u R^* > \Delta$, negative if $\tau_u R^* < \Delta$. Therefore, the distortion term is positive if $t_d > t_u$ and $\tau_u R^* > \Delta$ or $t_d < t_u$ and $\tau_u R^* < \Delta$. If the first-order condition of a function concave in k_u is greater than 0 at k_u^* , then the optimal value of k_u must be greater than k_u^* . Therefore, the firm overinvests in k_u relative to the benchmark.

If $t_d > t_u$ and $\tau_u R^* < \Delta$ or $t_d < t_u$ and $\tau_u R^* > \Delta$, then the second term is negative. Hence, the optimal k_u is less than k_u^* .

The first-order condition for k_d is:

$$-1 - \sum_{n=1}^N D_n \gamma_n v_d'(k_d) \tau_d + \sum_{n=1}^N D_n \delta_n t_d \doteq 0,$$

satisfied at $k_d = k_d^*$. Therefore, the regulation induces no distortion in k_d . ■

Proof of Proposition 2

The introduction of *ex post* discretion has no effect on the second-order conditions. Therefore, the function is still concave in k_u and k_d . If $t_d > t_u$, then the distortion term in the first derivative is:

$$(t_d - t_u) \left[(R^* + \mu\Phi) + \sum_{n=1}^N D_n [\gamma_n v_u'(k_u) + \delta_n] \right].$$

If $t_u > t_d$, then it is instead:

$$(t_d - t_u) \left[(R^* - \mu\Phi) + \sum_{n=1}^N D_n [\gamma_n v_u'(k_u) + \delta_n] \right].$$

The derivative becomes either more positive or less negative as Φ increases. Therefore, if

the regulation induces overinvestment in k_u , then the overinvestment is increasing in Φ , and if the regulation induces underinvestment, then the underinvestment is decreasing in Φ . ■

Proof of Lemma 3

The derivative of the present value of the transfer payments with respect to k_d evaluated at $k_d = k_d^*$ (the substitution in line 2) is:

$$\begin{aligned} \frac{\partial}{\partial k_d} \left[\sum_{n=1}^N D_n \gamma_n (p - v_d(k_d)) - \sum_{n=1}^N D_n \delta_n k_d - R^* k_d \right] &= -R^* - \sum_{n=1}^N D_n \delta_n - \sum_{n=1}^N D_n \gamma_n v'_d(k_d) \\ &= -R^* + \frac{1}{\tau_d} \left[1 - \sum_{n=1}^N D_n \delta_n \right]. \end{aligned}$$

Therefore the derivative is positive (negative) if $\tau_d R^* < \Delta$ ($\tau_d R^* > \Delta$).

The proof of the special case is symmetric to the proof of Lemma 2 and is omitted. ■

Proof of Proposition 3

Now the objective is Equation (11). $\Pi_{uu} = -\sum_{n=1}^N D_n \gamma_n \tau_d v''_u(k_u) < 0$, $\Pi_{dd} = -\sum_{n=1}^N D_n \gamma_n \tau_u v''_d(k_d) < 0$, and $\Pi_{ud} = \Pi_{du} = 0$. Therefore, the function is concave. The first derivative with respect to k_d is:

$$\left[-1 - \sum_{n=1}^N D_n \gamma_n v'_d(k_d) \tau_d + \sum_{n=1}^N D_n \delta_n t_d \right] + (t_d - t_u) \left[-R^* - \sum_{n=1}^N D_n [\gamma_n v'_d(k_d) + \delta_n] \right].$$

At k_d^* , the first term in the large brackets is 0. The sign of the distortion term (the second term) depends on both the relative tax rates and the sign of the last term in large brackets. As Lemma 3 establishes, the last term is negative at k_d^* if $\tau_d R^* > \Delta$, positive if $\tau_d R^* < \Delta$. Therefore, the distortion term is positive if $t_d > t_u$ and $\tau_d R^* < \Delta$ or $t_d < t_u$ and $\tau_d R^* > \Delta$. If the first-order condition of a function concave in k_d is greater than 0 at k_d^* , then the optimal value of k_d must be greater than k_d^* . Therefore, the firm overinvests in k_d relative to the benchmark.

If $t_d > t_u$ and $\tau_d R^* > \Delta$ or $t_d < t_u$ and $\tau_d R^* < \Delta$, then the second term is negative. Hence, the optimal k_d is less than k_d^* .

The first-order condition for k_u is:

$$-1 - \sum_{n=1}^N D_n \gamma_n v'_u(k_u) \tau_u + \sum_{n=1}^N D_n \delta_n t_u \doteq 0,$$

satisfied at $k_u = k_u^*$. Therefore, the regulation induces no distortion in k_u . ■

Proof of Proposition 4

The introduction of *ex post* discretion has no effect on the second-order conditions. Therefore, the function is still concave in k_u and k_d . If $t_d > t_u$, then the distortion term in the first derivative is:

$$(t_d - t_u) \left[-(R^* - \mu\Phi) - \sum_{n=1}^N D_n [\gamma_n v'_d(k_d) + \delta_n] \right].$$

If $t_u > t_d$, then it is instead:

$$(t_d - t_u) \left[-(R^* + \mu\Phi) - \sum_{n=1}^N D_n[\gamma_n v'_d(k_d) + \delta_n] \right].$$

The derivative becomes either more positive or less negative as Φ increases. Therefore, if the regulation induces overinvestment in k_d , then the overinvestment is increasing in Φ , and if the regulation induces underinvestment, then the underinvestment is decreasing in Φ . ■

Proof of Proposition 5

The objective function is Equation (17). $\Pi_{uu} = -\frac{1}{2} \sum_{n=1}^N D_n \gamma_n v''_u(k_u)(\tau_d + \tau_u) < 0$, $\Pi_{dd} = -\frac{1}{2} \sum_{n=1}^N D_n \gamma_n v''_d(k_d)(\tau_d + \tau_u) < 0$, and $\Pi_{du} = \Pi_{ud} = 0$. Therefore the function is concave. The derivative with respect to k_u is:

$$\left[-1 - \sum_{n=1}^N D_n \gamma_n v'_u(k_u) \tau_u + \sum_{n=1}^N D_n \delta_n t_u \right] + \frac{1}{2} (t_d - t_u) \sum_{n=1}^N D_n \gamma_n v'_u(k_u).$$

At k_u^* , the term in brackets is 0. If $t_u > t_d$, then the second term is positive. If the first-order condition of a function concave in k_u is greater than 0 at k_u^* , then the optimal value of k_u must be greater than k_u^* . Therefore, the firm overinvests in k_u relative to the benchmark. If $t_d > t_u$, then the second term is negative. Hence, the optimal k_u is less than k_u^* .

The derivative of the objective with respect to k_d is:

$$\left[-1 - \sum_{n=1}^N D_n \gamma_n v'_d(k_d) \tau_d + \sum_{n=1}^N D_n \delta_n t_d \right] - \frac{1}{2} (t_d - t_u) \sum_{n=1}^N D_n \gamma_n v'_d(k_d).$$

At k_d^* , the term in brackets is 0. If $t_d > t_u$, then the second term is positive. If the first-order condition of a function concave in k_d is greater than 0 at k_d^* , then the optimal value of k_d must be greater than k_d^* . Therefore, the firm overinvests in k_d relative to the benchmark. If $t_u > t_d$, then the second term is negative. Hence, the optimal k_d is less than k_d^* . ■

Proof of Proposition 6

With the addition of *ex post* income shifting, $\Pi_{uu} = -\frac{1}{2} \sum_{n=1}^N D_n \gamma_n v''_u(k_u)[(1 + \phi)\tau_d + (1 - \phi)\tau_u]$ and $\Pi_{dd} = -\frac{1}{2} \sum_{n=1}^N D_n \gamma_n v''_d(k_d)[(1 + \phi)\tau_u + (1 - \phi)\tau_d]$. These are both negative as long as $\phi < 1$. Therefore the function is still concave. Revisiting the derivatives with respect to k_u and k_d :

$$\left[-1 - \sum_{n=1}^N D_n \gamma_n v'_u(k_u) \tau_u + \sum_{n=1}^N D_n \delta_n t_u \right] + \frac{1}{2} (1 + \phi)(t_d - t_u) \sum_{n=1}^N D_n \gamma_n v'_u(k_u)$$

and

$$\left[-1 - \sum_{n=1}^N D_n \gamma_n v'_d(k_d) \tau_d + \sum_{n=1}^N D_n \delta_n t_d \right] - \frac{1}{2} (1 + \phi)(t_d - t_u) \sum_{n=1}^N D_n \gamma_n v'_d(k_d).$$

If $t_d > t_u$, $\phi = \Phi$, then the distortion term for k_u becomes more negative as Φ increases, and the distortion term for k_d becomes more positive as Φ increases. Therefore, the underinvestment in k_u and overinvestment in k_d are both more severe. If $t_d < t_u$, $\phi = -\Phi$, the

distortion term for k_u becomes less positive as Φ increases, and the distortion term for k_d becomes less negative as Φ increases. Therefore, both the overinvestment in k_u and the underinvestment in k_d are less severe. ■

REFERENCES

- Brealey, R., and S. Myers. 1996. *Principles of Corporate Finance*. New York, NY: McGraw-Hill.
- Graham, G. 1992. Clinton sets sights on foreign companies. *The Financial Times* (June 24): 6.
- Graham, J., and C. Harvey. 2001. The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics* 60 (May): 187–243.
- Halperin, R., and B. Srinidhi. 1987. The effects of the U.S. income tax regulations' transfer pricing rules on allocative efficiency. *The Accounting Review* 62 (October): 686–707.
- , and ———. 1996. U.S. income tax transfer pricing rules for intangibles as approximations of arm's length pricing. *The Accounting Review* 71 (January): 61–80.
- Harris, D., and R. Sansing. 1998. Distortions caused by the use of arm's length transfer prices. *The Journal of the American Taxation Association* 20: 40–50.
- Judis, J. 1993. A great American rip-off: Phony pricing helps foreign firms evade U.S. taxes. *San Diego Union Tribune* (August 18): B7.
- Kant, C. 1988. Endogenous transfer pricing and the effects of uncertain regulation. *Journal of International Economics* 24 (2): 147–157.
- Samuelson, L. 1982. The multinational firm with arm's length transfer price limits. *Journal of International Economics* 13 (4): 365–374.
- Sandmo, A. 1974. Investment incentives and the corporate income tax. *The Journal of Political Economy* 82 (2): 287–302.
- Sansing, R. 1999. Relationship-specific investments and the transfer-pricing paradox. *Review of Accounting Studies* 4 (June): 119–134.
- Williamson, O. 1985. *The Economic Institutions of Capitalism*. New York, NY: The Free Press.

The Effects of Decision-Aid Use and Reliability on Jurors' Evaluations of Auditor Liability

D. Jordan Lowe

University of Nevada, Las Vegas

Philip M. J. Reckers

Stacey M. Whitecotton

Arizona State University

ABSTRACT: This study provides evidence on how auditors' use of decision aids affects jurors' evaluation of auditor legal liability, based on an experiment in which actual jurors responded to a hypothetical audit lawsuit. The results suggest that decision aids can have positive, negative, or neutral effects on auditors' legal liability, depending on how auditors use the decision aid and the reliability of the decision aid. For high-reliability aids, jurors attributed more responsibility for an audit failure to the auditor when the auditor overrode the recommendation of a decision aid than when the auditor did not use the decision aid. However, jurors attributed lower responsibility to an auditor who relied on the recommendation of a highly reliable decision aid, even though the aid turned out to be incorrect. In contrast to the high-reliability conditions, auditors' use of the decision aid had virtually no impact on jurors' liability judgments when the reliability of the decision aid was low.

Keywords: *decision aids, auditor liability, fraud detection.*

Data Availability: *Data will be made available by the authors upon request.*

I. INTRODUCTION

One of the biggest challenges facing the accounting profession is the threat of litigation due to audit failure. Although recent court rulings have brought some relief to the profession by slowing or reversing the trend toward expanded auditor liability (Siliciano 1997; Pacini et al. 2000), Cloyd et al. (1998, 130) conclude that "the issue is

The authors acknowledge the helpful comments and suggestions of two anonymous reviewers, Steve Kachelmeier (associate editor), Steve Kaplan, Paul Carlson, Ed O'Donnell, workshop participants at the University of Arizona and Arizona State University, and participants of the 2000 Accounting, Behavior, and Organizations Research Conference.

*Submitted September 1999
Accepted May 2001*

far from being resolved, and the uncertainty surrounding it creates a need for a continued flow of independent, unbiased evidence from academe."

In addition to litigation concerns, audit firms face changing market pressures, including increased competition (e.g., the elimination of bans on competitive bidding and advertising in the Code of Professional Conduct), lower audit fees, and the need to offer a dynamic mix of services in response to changing customer demands (Elliott 1998; Melancon 1998). To survive and thrive in this environment, audit firms must continually improve the efficiency and effectiveness of the audit process.

One approach that audit firms have taken to achieve these objectives is to incorporate technology and decision aids (Abdolmohammadi and Usoff 2000). Decision aids have the potential to improve audit quality by increasing the likelihood that auditors will detect and report financial statement misstatements. In addition, decision aids can increase judgment consensus, or the extent to which other auditors would make the same decision under similar circumstances (Ashton and Willingham 1989; Messier 1995). Consensus is particularly important for audit firms defending against claims of negligence, because jurors may evaluate auditors based on the standard of "what other actors (auditors) would do in the same case situation" (Kelley and Michela 1980, 461). Finally, the enhanced documentation that decision aids provide may prove useful in defense of an audit lawsuit (Francis 1994; Lys and Watts 1994).

Despite these potential benefits, some audit firms may be reluctant to use decision aids because of uncertainty about how decision aids will affect their legal liability, and in particular, the fear that decision aids will leave an incriminating audit trail that will come back to haunt the auditor at trial. For example, Anderson et al. (1995, 38) note that decision aids create a veritable "legal minefield" in which "the seeds of subsequent failure will be highlighted." Messier (1995, 214) expressed similar concerns about the legal ramifications of decision aids:

If an auditing firm makes an expert system available to its auditors and an auditor chooses not to use it, will the auditor and firm be held liable if some related aspect of the audit is later found defective? Or, suppose an auditor overrides a decision aid's recommendation. Will this be viewed as evidence of a lack of due professional care? Finally, suppose that an expert system makes an incorrect decision. Who is liable and what standard should be used to measure the performance of the expert system?

Although these questions are important to the audit profession, surprisingly little research has attempted to address them. Furthermore, the few studies that have investigated the legal implications of decision aids have demonstrated only negative effects from their use (Jennings et al. 1993; Anderson et al. 1995). In other words, the conclusion from existing research would be that decision aids sometimes hurt, but never help, auditors' legal liability.

The current study sheds further light on these issues by investigating audit failures stemming from undetected fraud when (1) the auditor chooses not to use a fraud-prediction decision aid; (2) the auditor uses an aid, then chooses to ignore its prediction of fraud; and (3) the auditor chooses to rely on an aid that incorrectly predicts no fraud. We also include a control condition in which the audit firm has not implemented decision aids, and provide a separate manipulation of the reliability of the decision aid.

Our results suggest that the answers to the questions Messier (1995) posed about the legal liability issues surrounding decision aids not only depend on how the auditor uses the aid during the audit, but also on the reliability of the decision aid. For highly reliable decision aids, jurors attributed more responsibility to the auditor for an audit failure when

the auditor ignored a decision aid that predicted fraud, and less when the auditor relied on a decision aid that failed to predict fraud, compared to a control condition in which decision aids were not implemented. Interestingly, jurors' responsibility attributions were unrelated to decision-aid use for the low-reliability conditions, even though this level of reliability reflects the current "state of the art" for fraud prediction decision aids (Bell and Carcello 2000). Our results, in contrast to those from previous research, suggest that decision aids may be advantageous in some situations but disadvantageous in others, depending on how the auditor uses the decision aid during the audit *and* on the reliability of the decision aid.

In addition to providing more complete evidence about the circumstances that dictate whether decision aids will work to the auditor's benefit or detriment, we also add to the literature by investigating how jurors' responsibility evaluations translate into damage awards against audit firms. Our results confirm Kadous' (2000) conjecture that jurors may consider different factors to determine damage awards than they do to assess responsibility (including the audit firm's financial resources). Understanding how jurors' responsibility attributions translate into damages is particularly important in the context of management fraud, since lawsuits filed against independent auditors frequently involve allegations of management fraud, and these cases are primary contributors to large damage payments by audit firms (Palmrose 1987, 1999).

The remainder of this paper is organized as follows. Section II reviews previous research related to the legal ramifications of decision aids and develops the hypotheses tested in the current study. Section III provides an overview of our research method. In Section IV we present the results of our study. The final section discusses implications, limitations, and directions for future research.

II. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Previous Research on Fraud Detection and Decision Aids

In 1997, the Auditing Standards Board issued Statement on Auditing Standards 82 (SAS No. 82) to clarify auditors' responsibilities for detecting management fraud (AICPA 1997; Mancino 1997). Although the previous standard (SAS No. 53) required auditors to consider fraud in the overall evaluation of audit risk, it did not distinguish misstatements based on *intent*. Because audit procedures designed to detect unintentional misstatements (errors) may not be effective at detecting intentional misstatements (fraud), SAS No. 82 requires auditors to evaluate and document the risk of fraud on every audit to provide reasonable assurance that material misstatements are detected (Shibano 1990; Zimbelman 1997).

Even with the new standard in place, it may be difficult for auditors to detect management fraud. Compared to many routine aspects of the audit process, fraud detection is a very complex task. SAS No. 82 identifies more than 30 risk factors that auditors should consider in their evaluations of fraud, but gives little guidance concerning how auditors should combine these "red flags" into an overall assessment of fraud risk. In addition, because intentional errors (fraud) occur much less frequently than unintentional misstatements (errors) (Loebbecke et al. 1989), most auditors will not have had enough experience with fraud to develop elaborate knowledge structures (Bonner 1990). In combination, these findings suggest that fraud detection is a domain in which decision aids can improve the audit process.

Recent studies have investigated decision aids designed to facilitate fraud detection. Bell and Carcello (2000), for example, evaluated several logistic regression models that predict the likelihood of fraud based on the presence or absence of various risk factors identified in SAS No. 82. They found that their best model was more accurate in classifying

fraud and nonfraud cases than were practicing auditors, suggesting that audit firms consider using such models as decision aids to help them predict the likelihood of management fraud. Eining et al. (1997) found that this type of decision aid improved auditors' ability to discriminate between fraud and nonfraud cases and resulted in higher consensus among auditors. Consensus is important in a legal defense because jurors may judge auditors according to the standard of what other auditors would do in a similar situation.

From these and other studies, we submit the following observations: (1) the detection of management fraud is an important part of auditors' responsibilities; (2) fraud detection is a complex task with which auditors are likely to struggle due to lack of experience and ill-defined professional guidelines; and (3) although decision aids have the potential to improve the audit process, they also raise a number of complex legal issues that are not yet fully understood. The next section explores these legal implications in more detail and develops hypotheses.

Legal Implications of Decision Aids

Kadous (2000, 339) conceptualizes the juror's task as involving three steps: (1) assessing standards of care, (2) assessing audit quality, and (3) comparing the two to determine auditor responsibility for alleged audit failure. Thus, to fully understand the legal implications of decision aids, we must consider both their effect on audit quality and whether they change the standard of care (i.e., whether they raise the bar) against which auditors will be evaluated. As previously discussed, decision aids have the *potential* to improve audit quality by increasing the likelihood that auditors will detect and report material misstatements. At the same time, decision aids could shift perceived standards of care by providing a ready benchmark for evaluating auditor performance. For example, Jennings et al. (1993) suggest that in the absence of explicit external standards for evaluating auditor performance, jurists may use the "internal guidance" provided by a decision aid as a surrogate standard against which to judge auditor performance.

The major concern for audit firms would be situations in which standards of care escalate without a corresponding increase in jurors' perceptions of audit quality. For example, failure to use an available decision aid might not affect *actual* audit quality, but could adversely affect jurors' *perceptions* about audit quality. Furthermore, the availability of a decision aid may make it salient to jurors (*ex post*) that the standard of care was not met, thus increasing auditor liability. On the upside, an audit firm could argue that using a decision aid provides sufficient evidence of due professional care, thereby reducing auditor liability.

Anderson et al. (1995) tested these propositions by asking judges how likely they would be to support damages against an audit firm under three decision-aid conditions: no decision aid used, full use of a decision aid, and partial use of a decision aid. As expected, judges were more likely to support damages against the auditor when the auditor only partially used a decision aid than when no decision aid was available. However, full use of the decision aid did not decrease judges' support for damages against the auditor.

Taken at face value, the results of Anderson et al. (1995) present a bleak outlook for audit decision aids, illustrating the negative consequences that can occur when decision aids are implemented but *not* fully used, without a positive benefit when aids are implemented *and* fully used. However, an important caveat in interpreting these results is that Anderson et al. (1995) simultaneously varied decision-aid use and the reliability, or "face validity," of the decision aid. More specifically, in the partial use condition, the decision aid generated 15 potential explanations for an unusual financial statement fluctuation (and the auditor investigated only 10), but the aid generated only 10 explanations in the full use condition (and the auditor investigated all of them). Anderson et al. (1995) acknowledge

that this confound may have contributed to the insignificant results in the full use condition, and conclude that full use of a lower-quality decision aid may have been insufficient to meet judges' due professional care standards. They suggest that "further research should consider alternative methods of conveying decision-aid competence," such as through probabilistic information obtained as part of the decision-aid validation process (Anderson et al. 1995, 52–53).

The current study disentangles these effects by separately manipulating decision-aid use and reliability. Building on Kadous (2000), we expect that jurors will determine whether the auditor should be held responsible for an audit failure by comparing audit quality to some standard of care. The extent to which the auditor uses a decision aid during the audit is likely to affect jurors' perceptions about audit quality.¹ On the other hand, we expect that jurors will only view a decision aid as an implied standard of care if they perceive that the decision aid is reliable. Thus, our general expectation is that decision-aid use will influence jurors' responsibility attributions only if the aid is reliable:

H1: Jurors' evaluations of auditor responsibility for audit failure will be influenced by the extent to which the auditor uses the decision aid, but only if the aid is reliable.

Next, we break this overall interactive hypothesis into three specific hypotheses (H1a, H1b, and H1c) that correspond to the decision aid scenarios described by Messier (1995). In each case, we compare jurors' responsibility evaluations to a control condition in which no decision aid was available to the auditor.

One of the scenarios Messier (1995, 214) described was a situation in which "an auditor overrides a decision aid's recommendation. Will this be viewed as evidence of a lack of due professional care?" We expect the answer to this question will depend on whether jurors believe the decision aid is reliable. Specifically, we expect that overriding the recommendation of a reliable decision aid will make it salient to jurors that the standard of care was not met, thereby increasing jurors' attribution of responsibility to the auditor:

H1a: Jurors will assess higher auditor responsibility for audit failure when the auditor overrides the recommendation of a decision aid than when no decision aid is available, but only if the aid is reliable.

The next situation Messier (1995, 214) described was one in which "an expert system makes an incorrect decision. Who is liable and what standard should be used to measure the performance of the expert system?" In other words, will the fact that the auditor followed the recommendation of a decision aid (even if incorrect) provide sufficient evidence that the standard of care was met? Once again, we believe the answer will depend on whether the aid is reliable:

H1b: Jurors will assess lower auditor responsibility for an audit failure when the auditor relies on an incorrect decision aid than when no decision aid is available, but only if the aid is reliable.

Finally, we examine the situation in which "an auditing firm makes an expert system available to an auditor and an auditor chooses not to use it" (Messier 1995, 214). Our expectation is that jurors will penalize auditors for failing to use a reliable decision aid:

¹ This is consistent with Kadous' (2000) definition of audit quality, which she operationalized as whether the auditor consulted a specialist regarding inventory measurement.

H1c: Jurors will assess higher auditor responsibility for audit failure when the auditor fails to utilize a decision aid than when no decision aid is available, but only if the aid is reliable.

Next, we investigate how jurors' responsibility assessments translate into damage awards against the auditor. In audit liability cases, jurors must determine whether the audit firm (defendant) is responsible for losses incurred by the claimant (plaintiff) that result from an audit failure. If jurors attribute the cause of the loss to the auditor's actions, as opposed to other potential causes such as the situation or the actions of the claimant, then they are likely to assess damages against the auditor. Thus, we expect jurors' damage awards to be positively related to their assessment of the auditor's responsibility for the audit failure. In addition, other factors may influence jurors' assessment of damages, independent of their assessment of responsibility. As Kadous (2000, 340) notes, "once a juror has determined that an auditor is financially responsible for a plaintiff's loss, the juror will use a second set of factors to determine the amount of the liability." She suggests that although these factors may overlap with the factors that jurors use to evaluate responsibility, other factors, such as the dollar value of the plaintiff's loss, the auditor's ability to pay, and the availability of additional defendants, may influence damage awards even if they do not influence evaluations of responsibility.

In this study, we investigate whether the auditor's ability to pay (proxied by audit firm size) is an additional determinant of damage awards. Although prior research has suggested that litigation against independent auditors is often motivated by larger firms' perceived "deep pockets," (Carcello and Palmrose 1994; Lennox 1999), previous research on the relation between audit firm size and audit litigation rates has produced mixed results (Palmrose 1988; Stice 1991; Bonner et al. 1998).² These prior studies investigated the relation between audit firm size and the *incidence* of litigation, as opposed to the eventual *outcome* of the litigation.³ The incidence of litigation against large audit firms likely depends on more than their perceived deep pockets. For instance, large audit firms arguably perform a higher quality audit and have a greater incentive to avoid litigation than small firms (Dye 1993; Latham and Linville 1998). These factors would tend to mitigate the effect of large audit firms' deep pockets on the *incidence* of litigation.

To our knowledge, prior research has not investigated the effect of audit firm size on the amount of damages jurors award for audit failures. Once jurors have determined an outcome, they consider the damages that should be assessed against the defendant audit firm to compensate the injured party. At that point, the firm's ability to pay may become an important factor in the award of damages against the audit firm (Kadous 2000). This may be particularly true for jurors (as compared to judges), because jurors may be more prone to bias from extraneous factors, such as the defendant's resources (Palmrose 1991; Cloyd et al. 1998). This reasoning leads to the following hypotheses:

² The Private Securities Litigation Reform Act of 1995 replaced the long standing "joint and several" liability rule with a modified "proportionate" liability rule for most federal cases (King and Schwartz 1997; Hillegeist 1999). Although this change should reduce the likelihood that auditors and other "deep pocket" defendants will bear a disproportionate share of the damages from audit failure, it is not likely to eliminate the issue entirely. For example, because "joint and several" liability still applies in many state jurisdictions, plaintiffs may simply shift their litigation from federal courts to more favorable state venues (Palmrose 1997). Furthermore, Cloyd et al. (1998) note that jurors may circumvent the "proportionate" liability rule by apportioning a greater share of the liability to "deep pocket" defendants to compensate for the limited recoveries expected from "shallow pocket" defendants.

³ These studies assume that plaintiff attorneys anticipate how juries will attribute responsibility and assess damages in audit trials, and factor these considerations into their decision to initiate litigation against audit firms.

H2: Jurors' damage awards will be positively related to their assessment of the auditor's responsibility for the audit failure.

H3: Jurors will assess higher damage awards against large audit firms than against small audit firms.

III. METHOD

Participants

We conducted an experiment with 149 jurors who were waiting to be called for jury service.⁴ Although most audit litigation cases are settled out of court, those that make it to trial are more often decided by jury than by judge, and typically have a lower auditor success rate than judge trials (Palmrose 1991). Jurors completed the case instrument in quiet conference rooms and were subsequently compensated \$5.00. Sixty-nine (46.3 percent) of the participants were male and 80 (53.7 percent) were female. The pool of jurors was diverse, with an average age of 40.6 (standard deviation = 13.5), and a wide range of income and education levels.

Task

Participants read and answered questions about a hypothetical audit lawsuit in which an investor was suing a public accounting firm for alleged audit failure. We used separate envelopes for each part of the task to prevent jurors from looking back. The case instrument began with an overview of the audit process including a description of Generally Accepted Auditing Standards. We also explained the auditors' responsibilities for assessing client risk and their overall concern for management fraud. This introductory material educated jurors (as would be done in an actual trial) prior to distribution of information specific to the case.

The case focused on the audit of a publicly traded client in the medical technology industry that had recently experienced a favorable earnings trend. We described the audit firm and the approach the firm and audit manager took in using decision aids in the conduct of the audit (described in the "Design" section below). After performing the audit, the audit firm concluded that the potential for fraud was low and therefore did not perform additional audit procedures. The firm subsequently issued an unqualified audit opinion. Based on the positive trend in earnings in the audited financial statements, an investor purchased 200,000 shares of the firm's stock. It was later revealed that the client had fraudulently overstated net income for the last few years, and that in reality the company was barely breaking even. This discovery caused the client's stock price to drop and the investor (plaintiff) who had relied on the audited financial statements brought a lawsuit against the audit firm. The audit firm claimed that the audit was planned and executed with due professional care and that the court should not hold the firm responsible for failing to detect the fraud.

Design

We used a $4 \times 2 \times 2$ design, with the following between-subjects variables: decision-aid use (four levels), decision-aid reliability (two levels), and audit firm size (two levels).

⁴ Prior arrangements had been made with the court administrator. Jurors are called from a general jury list to serve in a juror pool. The jury list includes names drawn from county driver's license registration files and from a master file of registered voters from the Board of Elections. Individuals are deleted from the jury list if they (1) are less than 18 years old, (2) are not current residents of the county, or (3) are presently exempt from jury duty. The court administrator sorts the random file into numerical order and each jury pool is selected sequentially. Plaintiff or defense counsel can then dismiss jurors due to perceived conflicts or biases (i.e., occupation, attitudes, etc.).

The decision-aid use conditions included a control group in which the firm had not implemented decision aids, plus three conditions designed to represent the scenarios described by Messier (1995) in which an auditor (1) chose to override the recommendation of the decision aid, (2) chose to rely on the recommendation of an incorrect decision aid, or (3) chose not to use an available decision aid during the audit.⁵ The decision aid described in the case was similar in structure to the fraud prediction aid proposed by Bell and Carcello (2000).

We described the reliability of the decision aid as either an 81 percent (low) or a 90 percent (high) historical accuracy rate. We chose the 81 percent rate based on previous research on fraud prediction decision aids.⁶ For example, Boatsman et al. (1997) reported that their fraud prediction decision aid correctly classified 81 percent of fraud and nonfraud cases. Bell and Carcello (2000) reported that their best model correctly classified 30/37 (81 percent) of fraud cases and 123/143 (86 percent) of nonfraud cases. Once we chose 81 percent as a "realistic" hit rate, our next step was to choose a rate that jurors would view as more/less reliable. After pilot testing with student participants, we chose to go with a higher level of accuracy because we did not believe that jurors would find an 81 percent aid to be highly reliable. Research in psychology (Arkes et al. 1986) and accounting (Ashton 1990; Whitecotton 1996) suggests that decision makers often mistakenly believe they can outperform a statistically superior decision aid. For example, Boatsman et al. (1997) found that auditors were reluctant to rely on an 81 percent accurate fraud prediction aid, even though it outperformed unaided auditors (59 percent) and auditors were told that the aid would "outperform even firm experts." We expected that jurors would be even less likely to perceive an 81 percent aid as reliable because they would be evaluating the aid's performance in hindsight. Based on this reasoning, we chose 90 percent for the high-reliability condition.

The third independent variable was audit firm size, which we described as either "a large international public accounting firm with offices in more than 60 cities, with nearly 20,000 professionals" or "a relatively small auditing firm with offices in four cities and three states, with about 230 professionals."

Two aspects of our design warrant further discussion. Although we made these design choices in an effort to provide a complete and realistic experimental design, they create limitations that should be kept in mind when interpreting our results. First, because the auditor always concluded that the potential for fraud was low when fraud was actually present (otherwise there would be no lawsuit), the decision aid's prediction had to vary across the decision-aid use conditions. For instance, in those conditions in which the auditor did not rely on the decision aid, the case indicated that the decision aid would have predicted fraud. However, when the auditor chose to rely on the decision aid (which proved incorrect), the aid predicted no fraud. Thus, we cannot say for certain whether any differences in juror responses across these scenarios are due to our intended manipulation of decision-aid use, or to the prediction of the decision aid itself (fraud/no fraud). However, manipulation checks included in the exit questionnaire suggest that jurors paid more attention to whether

⁵ Based on current practice, we believed that auditors who would not use a computer-based decision aid would at least rely on some sort of checklist. To make the case as realistic as possible, we included a checklist when the auditor did not use a computer-based decision aid. We also acknowledge that checklists are a type of decision aid.

⁶ Most prior research has relied on the sample of 77 fraud and 305 nonfraud cases from Loebbecke et al. (1989) and Bell et al. (1991), respectively. Although the accuracy rates of the models vary depending on the red flags included in the model, the actual outcome (fraud or no fraud), and the cutoff criterion for predicting fraud/no fraud (Bell and Carcello 2000), 81 percent emerges as a "realistic" hit rate for fraud prediction decision aids.

the auditor used a decision aid during the audit than whether the aid predicted fraud or no fraud.⁷

Second, to create a fully crossed design, we included some description of decision-aid reliability in all conditions, including the “no decision aid” control group. To accomplish this, we added a statement from the plaintiff indicating that if the audit firm had used a decision aid that was available in the commercial market or used by other CPA firms (“packages that are correct about 90 percent (or 81 percent) of the time”), then it would have led to the discovery of fraud. Thus, although the audit firm had chosen not to implement decision aids, some form of decision aid was at least potentially available even in the control group.⁸

Dependent Variables

Our first dependent variable captures jurors’ beliefs about the extent to which they should hold the audit firm responsible for the losses that the plaintiff claims resulted from an audit failure. We asked three questions to measure these beliefs:

- A. Do you feel that the audit manager (and the audit firm) made the right decision in concluding that fraud was not present and that there was no need to perform additional audit procedures?
- B. How competent did you perceive the audit manager (and the audit firm) to be in performing its duties in the audit of this client?
- C. To what extent do you believe that the plaintiff must assume normal investment risks when purchasing stock, and therefore is largely responsible for their own loss?

Participants responded to these questions on a ten-point Likert scale. The questions were reverse-coded so that higher (lower) values suggest that the auditor was more (less) responsible for the audit failure. We then used factor analysis to combine these three questions into an overall measure of juror attribution of responsibility.⁹ Factor analysis attempts to aggregate similar, individual factors into general, underlying constructs. Using principle components analysis, the three questions loaded on a single factor, producing an eigenvalue of 1.8 and a Cronbach’s alpha of 0.66. The resulting factor scores serve as the dependent measure for jurors’ assessments of auditor responsibility. These scores have a mean of 0 and a variance of 1. Thus, negative scores indicate that jurors attributed “lower than average” responsibility for the audit failure to the auditor, and positive scores imply “above average” responsibility for the audit failure.

The second dependent measure relates to jurors’ assessments of damages against the audit firm. We asked jurors to respond to the following three questions using a ten-point Likert scale:

⁷ When asked whether the audit firm used a decision aid during the audit, 85 percent of participants responded correctly, whereas only 56 percent responded correctly to a similar question asking whether the decision aid predicted fraud or no fraud.

⁸ Although we recognize that this has the potential of weakening our results, we felt it was important to provide a control condition in our design. Furthermore, Sutton et al. (1995) note that audit firms are increasingly called upon to provide this type of expert testimony on behalf of plaintiffs who have filed lawsuits against other audit firms.

⁹ We also conducted two alternative analyses. First, we performed a multivariate analysis (MANOVA) with each of the individual questions treated as dependent variables. We also combined the individual questions into an aggregate measure using an equally weighted scoring rule (Kadous 2000). The results from each analysis were qualitatively similar to the factor analysis results.

- A. To what extent do you believe that the CPA firm's audit was ineffective, and therefore the CPA firm should reimburse the plaintiff?
- B. What is the likelihood that you would support the plaintiff's call for *some* amount of damages by the auditor?
- C. What is the likelihood that you would support the plaintiff's call for the *total* amount of damages by the auditor?

Similar to the procedure described above, we used factor analysis to combine these three questions into an overall measure of jurors' assessments of damages. The three questions loaded on a single factor, producing an eigenvalue of 2.1 and a Cronbach's alpha of 0.79. The resulting factor scores, which have a mean of 0 and a variance of 1, comprised the dependent measure of juror assessment of damages against the auditor. Thus, positive values reflect above-average damage assessments, and vice versa.

Finally, we collected the following demographic variables that could influence jurors' assessments of responsibility and/or damages: gender, age, education, and income level (see Hastie 1993).¹⁰

IV. RESULTS

Responsibility Evaluations (H1)

The overall prediction of H1 was that auditors' use of decision aids would influence juror responsibility attributions only if the decision aid is reliable. We test this interactive hypothesis using an ANOVA model, with the experimental manipulations of decision-aid use (four levels) and reliability (two levels) serving as the independent variables, and the responsibility factor score designated as the dependent variable. Preliminary analysis revealed that the other model variables—audit firm size and juror demographics—were not significantly related to the dependent variable. Therefore, we excluded these variables from the analyses related to H1. The ANOVA results appear in Panel A of Table 1; the means and standard deviations for the dependent measure are shown in Panel B.

As shown in Panel A of Table 1, the interaction between decision-aid use and decision-aid reliability is statistically significant ($F = 2.82$; $p = 0.041$), supporting H1. To provide more evidence on the specific form of this overall interaction, we performed planned comparisons to test the predictions of H1a, H1b, and H1c. Figures 1, 2, and 3 display the relevant cell means for each of these hypotheses in graphical format.

Hypothesis 1a predicted that jurors would attribute more responsibility to an auditor who chose to override the aid than if no decision aid was available, but only if the decision aid is reliable. The cell means displayed in Figure 1 are consistent with the form of this predicted interaction. Planned comparisons confirm that jurors held auditors more responsible for overriding the recommendation of the decision aid (compared to the control condition) when the reliability of the decision aid was high ($t = 2.61$, $p = 0.007$), but not when the reliability of the decision aid was low ($t = 0.28$, $p = 0.777$). Thus, H1a is supported.

Our next hypothesis (H1b) relates to the condition in which the auditor relied on the recommendation of the decision aid, but the aid was incorrect. We expected that jurors would assess lower responsibility for the audit failure when the auditor followed the decision aid's recommendation compared to the control group in which the aid was not available, but only if the aid is reliable. The cell means displayed in Figure 2 are consistent with the form of this predicted interaction. Planned comparisons show that jurors attributed

¹⁰ We also included these demographic variables to check on the randomization of the various levels of the independent variable. Statistical tests performed on the demographic variables indicate randomization was successful across treatment levels.

TABLE 1
The Effects of Decision-Aid Use and Decision-Aid Reliability (H1)
On Jurors' Responsibility Assessments^a

Panel A: Analysis of Variance Model

<i>Source of Variation</i>	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p-value</i>
Overall Model	19.06	7	2.72	2.98	0.006
Decision Aid Use ^b	5.21	3	1.74	1.90	0.133
Decision Aid Reliability ^c	6.59	1	6.59	7.21	0.008
Use × Reliability	7.75	3	2.58	2.82	0.041

Panel B: Mean Juror Attribution Score, (Standard Deviation), Cell Sizes

	<i>Low-Reliability Decision Aid</i>	<i>High-Reliability Decision Aid</i>
Control Group— No Decision Aid	−0.324 (0.942) n = 17	0.128 (0.710) n = 22
Auditor Overrides the Decision Aid	−0.236 (0.924) n = 20	0.816 (0.958) n = 18
Auditor Relied on Incorrect Decision Aid	−0.093 (1.204) n = 18	−0.336 (0.925) n = 18
Decision Aid Available, But Not Used	−0.201 (1.005) n = 18	0.225 (0.970) n = 18

^a The dependent variable was obtained through principal component analysis on three individual questions related to jurors' attributions of auditor responsibility. This method combines the individual measures into a standardized factor score that, by construction, has a mean of 0 and a variance of 1. Thus, negative scores indicate that jurors attributed "lower than average" responsibility for the audit failure to the auditor, and positive scores imply "above average" responsibility for the audit failure.

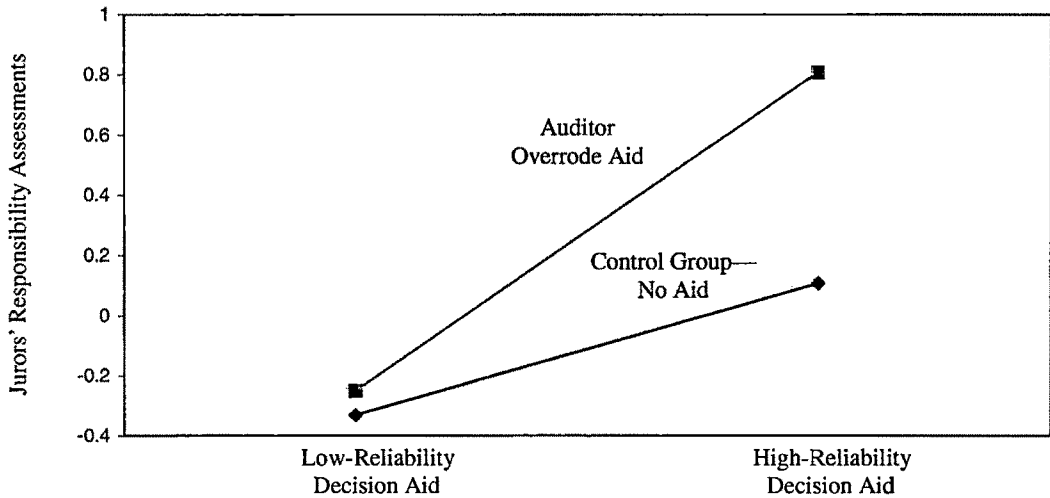
^b Decision aid use included (1) a control group in which the firm had not implemented decision aids, plus three conditions in which an auditor (2) chose to override the recommendation of the decision aid, (3) chose to rely on the recommendation of an incorrect decision aid, or (4) chose not to use an available decision aid during the audit.

^c The reliability of the decision aid was described as either 81 percent (low) or a 90 percent (high) historical accuracy rate.

lower responsibility to the auditor who relied on a highly reliable decision aid than the control group ($t = 1.80$, $p = 0.040$), but there was no significant difference for the low-reliability decision aid ($t = 0.63$, $p = 0.533$). Thus, H1b is supported.

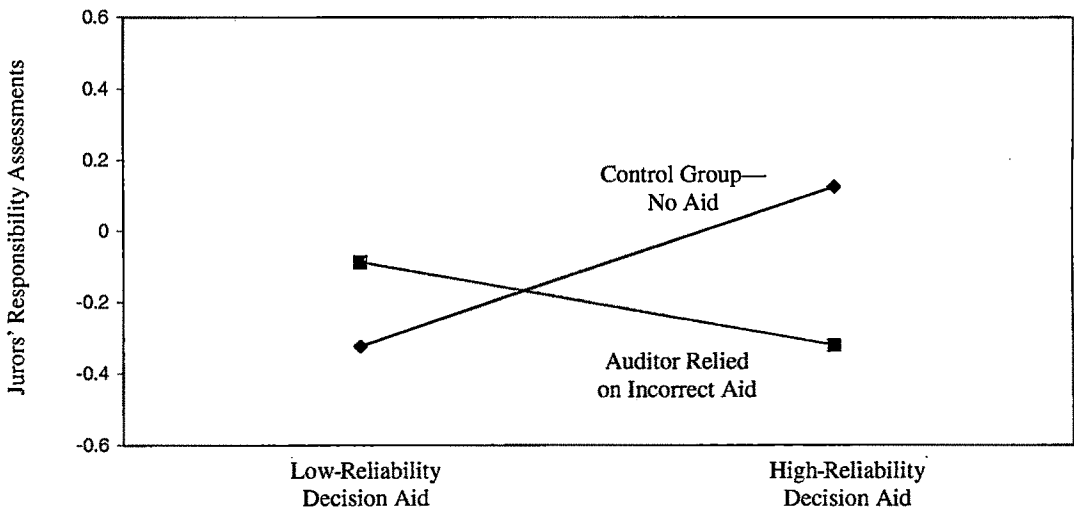
Finally, H1c relates to the condition in which the decision aid was available, but the auditor chose not to use it during the audit. Once again, our expectation was that failure to use an available decision aid would lead jurors to attribute more responsibility to the

FIGURE 1
Effects of the Auditor Overriding the Decision Aid and Decision-Aid Reliability on Jurors' Responsibility Assessments (H1a)



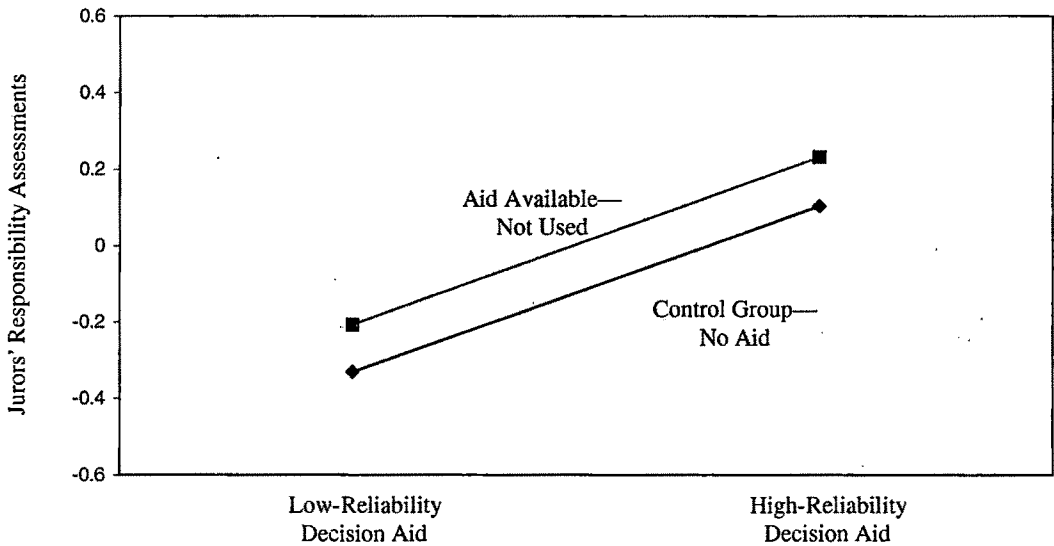
Scores above 0 indicate that jurors attributed "higher than average" responsibility to the auditor for the audit failure, and vice versa.

FIGURE 2
Effects of the Auditor Relying on an Incorrect Decision Aid and Decision-Aid Reliability on Jurors' Responsibility Assessments (H1b)



Scores above 0 indicate that jurors attributed "higher than average" responsibility to the auditor for the audit failure, and vice versa.

FIGURE 3
Effects of the Auditor Not Using an Available Decision Aid and Decision-Aid Reliability on Jurors' Responsibility Assessments (H1c)



Scores above 0 indicate that jurors attributed “higher than average” responsibility to the auditor for the audit failure, and vice versa.

auditor than in the control condition, but only when the aid was reliable. Figure 3 does not support this predicted interaction. Planned *t*-tests show no significant differences in jurors' mean responsibility evaluations between the condition in which the decision aid was available (but not used) and the control group, regardless of whether decision-aid reliability was low ($t = 0.37$, $p = 0.713$) or high ($t = 0.36$, $p = 0.719$). Thus, H1c is not supported.

Although not hypothesized, the results displayed in Figure 3 suggest that jurors attributed more responsibility to the audit firm in the presence of a high-reliability decision aid than a low-reliability aid, regardless of the decision-aid condition ($t = 2.09$; $p = 0.039$). The only difference between these two conditions was how readily available the decision aid was for the auditor's use. In the control group, the audit firm had not implemented decision aids, but jurors were told that decision aids were commercially available and used by other CPA firms. In the treatment condition, the audit firm had implemented the decision aid, but the auditor chose not to use it during the audit in question. The results shown in Figure 3 suggest that jurors did not differentiate between these two conditions—jurors appear to penalize auditors for failing to use a high-reliability decision aid regardless of whether the decision aid was available within the audit firm or on the commercial market.

Damage Award Assessments (H2 and H3)

Next, we investigate whether jurors' responsibility evaluations (H2) and audit firm size (H3) influence jurors' award of damages against the audit firm. Table 2 reports the results of a regression model to test these hypotheses. The dependent measure was the combined factor score on the three damage-related questions, where positive scores indicate “higher

TABLE 2

Results of Regressing Jurors' Damage Assessments on Jurors' Responsibility Evaluations (H2) and Audit Firm Size (H3)^a

<i>Variable</i>	<i>Coefficient</i>	<i>Standard Error</i>	<i>t-statistic</i>	<i>p-value</i>
Intercept	0.275	0.385	0.71	0.476
Responsibility Evaluation ^b	0.598	0.063	9.58	<0.001
Audit Firm Size ^c	0.362	0.124	2.91	0.004
Juror Age	0.012	0.005	2.52	0.013
Juror Gender ^d	-0.261	0.128	-2.04	0.043
Juror Education	-0.137	0.037	-3.76	<0.001
Juror Income	0.038	0.058	0.66	0.510

Adj. $R^2 = 0.442$

^a The dependent variable was obtained through principal component analysis on three individual questions related to jurors' assessments of damages. Higher (lower) values suggest that jurors were more (less) likely to award damages against the audit firm.

^b This variable was obtained through principal component analysis on three individual questions related to jurors' attributions of auditor responsibility. Higher (lower) values indicate that jurors attributed more (less) responsibility for the audit failure to the auditor.

^c Audit firm size was coded as either a 0 for "a relatively small auditing firm with offices in four cities and three states, with about 230 professionals" or 1 for "a large international public accounting firm with offices in more than 60 cities, with nearly 20,000 professionals."

^d Female jurors were coded 1 and male jurors were coded 0.

than average" damages against the audit firm, and vice versa. In addition to the hypothesized independent variables, the model includes juror age, gender, education, and income level as control variables.¹¹

The overall regression model is highly significant ($F = 20.51$; $p < 0.001$), with the six variables combining to explain approximately 44 percent of the variation in jurors' damage assessments (adjusted $R^2 = 0.442$). As H2 predicts, the relationship between jurors' responsibility evaluations and damage assessments was significantly positive ($t = 9.58$; $p < 0.001$), suggesting that jurors who assessed higher auditor responsibility for the audit failure were more likely to support damages against the auditor. The significantly positive coefficient for audit firm size ($t = 2.91$; $p = 0.004$) suggests that jurors were more likely to assess damages against a large audit firm than a small audit firm, supporting H3.

The regression model also shows significant effects for juror age ($t = 2.52$; $p = 0.013$), education level ($t = -3.76$; $p < 0.001$), and gender ($t = -2.04$; $p = 0.043$). These results suggest that older and less educated jurors were more likely to support damages against the auditor, and that male jurors were more likely to support damages against the audit firm than were female jurors. Juror income level was not significantly related to damage assessments ($t = 0.66$; $p = 0.510$).

¹¹ Since there were no significant interactions among the independent variables, only main effects were included in the model.

V. DISCUSSION AND IMPLICATIONS

This study adds to the limited evidence on the legal implications of decision aids by examining the effects of auditors' use of decision aids and decision-aid reliability on jurors' evaluations of auditor liability. Our results suggest that the extent to which auditors use decision aids during an audit influences jurors' legal liability judgments only if jurors perceive the decision aid as reliable. For example, jurors assessed higher responsibility for an audit failure when an auditor overrode the recommendation of a highly reliable decision aid than when the audit firm had not implemented decision aids. These results should be of interest to audit firms who are concerned that implementing decision aids may heighten their liability exposure.

On a more positive note, and in contrast to the results of Anderson et al. (1995), we found that jurors attributed *lower* responsibility to an auditor who followed the recommendation of a highly reliable decision aid, even though the aid turned out to be incorrect. However, regardless of decision-aid reliability, there were no differences in jurors' responsibility evaluations depending on whether (1) the decision aid was available within the firm but the auditor did not use it during that particular audit, and (2) the aid was available in the commercial market but had not been implemented by the audit firm. Apparently, jurors did not view these two scenarios as different enough to affect their legal liability judgments. Thus, our results suggest that highly reliable decision aids can have positive, negative, or neutral effects on auditors' legal liability, depending on how auditors use decision aids during the audit.

In contrast to the high-reliability condition, decision-aid use had virtually no effect on jurors' responsibility evaluations in the low-reliability condition. These insignificant results may be enlightening to audit firms, since this level of accuracy (81 percent) reflects the current "state of the art" for fraud prediction decision aids (Bell and Carcello 2000). Thus, the fear that some audit firms have expressed about the potential adverse legal consequences of decision aids may be unwarranted, at least for currently available fraud prediction decision aids. However, as model developers improve the predictive accuracy of such decision aids (by incorporating additional "red flags," for example), the legal liability stakes may increase.

Thus, audit firms appear to face a dilemma regarding the development and use of decision aids. If firms choose not to implement decision aids, then they may be disadvantaged by the lack of internal guidance and consistency these tools provide. In addition, by not implementing reliable decision aids, audit firms may be at a legal disadvantage if other firms have adopted decision aids, since auditors may be held responsible for failing to use a reliable decision aid, regardless of whether the aid was available internally or on the commercial market. However, if firms develop and use decision aids, then they may be constrained to adhere fully to the decision aids' recommendations, as these aids may serve as implied standards of performance in future litigation, particularly as decision aids gain reliability.

These results have important implications for audit firms' policies regarding the implementation and use of decision aids. For instance, should the firm's executive office mandate the use of decision aids in their audits? Although the results of our study suggest that such a policy may have advantages in a court of law, it may have some disadvantages in practice. Requiring auditors to use a decision aid may lead auditors to approach their tasks mechanistically (Glover et al. 1997), or, alternatively, may lead to dysfunctional behavior such as intentional circumvention of the aid (Kachelmeier and Messier 1990).

Another objective of this study was to investigate how jurors' responsibility evaluations translate into damage awards against the audit firm. Our results revealed that jurors' responsibility evaluations were significantly related to their damage assessments. This finding provides some evidence that jurors assess damages based upon the degree of corresponding responsibility, in contrast to concern that jurors provide damage assessments based strictly on nonmeritorious (frivolous) claims (Cloyd et al. 1996; Palmrose 1997).

Responsibility assessment was not the only factor affecting jurors' damage assessments. Jurors were also more likely to support damages against large audit firms than small audit firms, which provides further evidence of the perception of large audit firms as "deep pockets." Interestingly, although audit firm size and several juror demographic measures were related to jurors' damage assessments, these factors did not influence jurors' evaluations of auditor's responsibility for the audit failure. These results are consistent with Kadous' (2000) conjecture that jurors may use a different set of factors to determine damages than they do to assess responsibility. Our results suggest that although jurors were able to ignore factors such as the audit firm's "ability to pay" and their own personal beliefs when determining who was responsible for the audit failure, they did allow these factors to influence their damage awards. Additional research is necessary to understand this inconsistency and to determine how and why these factors influenced jurors' damage assessments in an audit liability scenario.

Our results also showed that jurors attributed less responsibility to the auditor when they perceived a decision aid as having low reliability, regardless of how the auditor had used the decision aid during the audit. Defense counsel could therefore emphasize the limitations and incomplete nature of decision aids. The results related to juror demographics may also prove useful to audit firms and their defense counsel in the selection of jurors.

Of course, one must view the results of this study in light of its limitations. In addition to the design limitations previously discussed, many of the items presented in the case may be nonfactual, ambiguous, or a source of contention in a court of law. In a more realistic setting, jurors would receive substantial judicial instructions along with plaintiff and defense counsels' opposing viewpoints. It is not clear how these factors would influence our findings.

Finally, we manipulated the reliability of the decision aid as either 81 percent (low) or 90 percent (high). Although our results suggest that jurors perceived these levels differently, we have no way of identifying the critical threshold for decision-aid reliability. It is also likely that other accuracy rates would produce different results. For example, jurors might penalize auditors for using an aid with a very low level of reliability. Future research should determine what jurors (and judges) consider the minimum accuracy level of decision aids before they are considered reliable and therefore relevant in auditor litigation.

REFERENCES

- Abdolmohammadi, M., and C. Usoff. 2000. A longitudinal study of applicable decision aids for detailed tasks in a financial audit. Working paper, Bentley College, Waltham, MA.
- American Institute of Certified Public Accountants (AICPA). 1997. *Consideration of Fraud in a Financial Statement Audit*. Statement on Auditing Standards No. 82. New York, NY: AICPA.
- Anderson, J. C., M. M. Jennings, S. E. Kaplan, and P. M. J. Reckers. 1995. The effects of using diagnostic decision aids for analytical procedures on judges' liability judgments. *Journal of Accounting and Public Policy* 14: 33–62.
- Arkes, H. R., R. M. Dawes, and C. Christensen. 1986. Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Performance* 37: 93–110.

- Ashton, R. H. 1990. Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback and justification. *Journal of Accounting Research* 28 (Supplement): 148–180.
- , and J. J. Willingham. 1989. Using and evaluating audit decision aids. In *Auditing Symposium IX: Proceedings of the 1988 Touche and Ross/University of Kansas Symposium on Auditing Problems*, edited by R. P. Srivastava, and J. E. Rebele, 1–25. Lawrence, KS: University of Kansas.
- Bell, T. B., and J. V. Carcello. 2000. A decision aid for assessing the likelihood of fraudulent financial reporting. *AUDITING: A Journal of Practice & Theory* 19 (Spring): 169–184.
- , S. Szykowny, and J. J. Willingham. 1991. Assessing the likelihood of fraudulent financial reporting: A cascaded logit approach. Working paper, KPMG LLP, Montvale, NJ.
- Boatsman, J. R., C. Moeckel, and B. Pei. 1997. The effect of decision consequences on auditors' reliance on decision aids in audit planning. *Organizational Behavior and Human Decision Processes* 71 (2): 211–247.
- Bonner, S. E. 1990. Experience effects in auditing: The role of task-specific knowledge. *The Accounting Review* 65 (January): 72–92.
- , Z-V. Palmrose, and S. M. Young. 1998. Fraud type and auditor litigation: An analysis of SEC accounting and auditing enforcement releases. *The Accounting Review* 73 (October): 503–532.
- Carcello, J. V., and Z-V. Palmrose. 1994. Auditor litigation and modified reporting on bankrupt clients. *Journal of Accounting Research* 32 (Supplement): 1–30.
- Cloyd, C. B., J. R. Frederickson, and J. W. Hill. 1996. Motivating factors in lawsuits against independent auditors: Experimental evidence on the importance of causality. *Journal of Accounting and Public Policy* 15 (Fall): 185–218.
- , ———, and ———. 1998. Independent auditor litigation: Recent events and related research. *Journal of Accounting and Public Policy* 17 (Summer): 121–142.
- Dye, R. A. 1993. Auditing standards, legal liability and auditor wealth. *Journal of Political Economy* 101 (October): 887–914.
- Eining, M. M., D. R. Jones, and J. K. Loebbecke. 1997. Reliance on decision aids: An examination of auditors' assessments of management fraud. *AUDITING: A Journal of Practice & Theory* 16 (Fall): 1–19.
- Elliott, R. K. 1998. Assurance services and the audit heritage. *The CPA Journal* 68 (June): 40–47.
- Francis, J. R. 1994. Auditing, hermeneutics, and subjectivity. *Accounting, Organizations and Society* 19 (April): 235–269.
- Glover, S. M., D. F. Prawitt, and B. C. Spilker. 1997. The influence of decision aids on user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational Behavior and Human Decision Processes* 72 (2): 232–255.
- Hastie, R. 1993. *Inside The Juror: The Psychology of Decision Making*. New York, NY: Cambridge University Press.
- Hillegeist, S. A. 1999. Financial reporting and auditing under alternative damage apportionment rules. *The Accounting Review* 74 (July): 347–369.
- Jennings, M., D. Kneer, and P. Reckers. 1993. The significance of audit decision aids on pre-case jurists' attitudes on perceptions of audit firm culpability and liability. *Contemporary Accounting Research* 9 (Spring): 489–507.
- Kachelmeier, S. J., and W. Messier. 1990. An investigation of the influence of a nonstatistical decision aid on auditor sample size decisions. *The Accounting Review* 65 (January): 209–226.
- Kadous, K. 2000. The effects of audit quality and consequence severity on juror evaluations of auditor responsibility for plaintiff losses. *The Accounting Review* 75 (July): 327–341.
- Kelley, H. H., and J. L. Michela. 1980. Attribution theory and research. In *Annual Review of Psychology*, edited by M. R. Rosenzweig, 457–501. Palo Alto, CA: Annual Reviews.
- King, R., and R. Schwartz. 1997. The Private Securities Litigation Reform Act of 1995: A discussion of three provisions. *Accounting Horizons* 11 (March): 92–106.
- Latham, C. K., and M. Linville. 1998. A review of the literature in audit litigation. *Journal of Accounting Literature* 17: 175–213.

- Lennox, C. S. 1999. Audit quality and auditor size: An evaluation of reputation and deep pockets hypotheses. *Journal of Business, Finance, and Accounting* 26 (Sept/Oct): 779–806.
- Loebbecke, J., M. Eining, and J. Willingham. 1989. Auditors' experience with material irregularities: Frequency, nature and detectability. *AUDITING: A Journal of Practice & Theory* 9 (Fall): 1–28.
- Lys, T., and R. L. Watts. 1994. Lawsuits against auditors. *Journal of Accounting Research* 32 (Supplement): 65–95.
- Mancino, J. 1997. The auditor and fraud. *Journal of Accountancy* 183 (April): 32–36.
- Melancon, B. C. 1998. The changing strategy for the profession, the CPA and the AICPA: What this means for the education community. *Accounting Horizons* 12 (December): 397–406.
- Messier, W. F. 1995. Research in and development of audit decision aid. In *Judgment and Decision-Making Research in Accounting and Auditing*, edited by R. H. Ashton, and A. H. Ashton, 207–228. New York, NY: Cambridge Press.
- Pacini, C., W. Hillison, and D. Sinason. 2000. Three's a crowd: An examination of state statutes and court decisions that narrow accountant liability to third parties for negligence. *Advances in Accounting* 17: 151–186.
- Palmrose, Z-V. 1987. Litigation and independent auditors: The role of business failures and management fraud. *AUDITING: A Journal of Practice & Theory* 6 (Spring): 90–103.
- . 1988. An analysis of auditor litigation and audit service quality. *The Accounting Review* 63 (January): 55–73.
- . 1991. Trials of legal disputes involving independent auditors: Some empirical evidence. *Journal of Accounting Research* 29 (Supplement): 149–185.
- . 1997. Audit litigation research: Do the merits matter? An assessment and directions for future research. *Journal of Accounting and Public Policy* 16 (Winter): 355–378.
- . 1999. *Empirical Research in Auditor Litigation: Considerations and Data*. Sarasota, FL: American Accounting Association.
- Shibano, T. 1990. Assessing audit risk from errors and irregularities. *Journal of Accounting Research* 28 (Supplement): 110–140.
- Siliciano, J. A. 1997. Trends in independent auditor liability: The emergence of a sane consensus? *Journal of Accounting and Public Policy* 16 (Winter): 339–353.
- Stice, J. D. 1991. Using financial and market information to identify pre-engagement factors associated with lawsuits against auditors. *The Accounting Review* 66 (July): 516–533.
- Sutton, S. G., R. Young, and P. McKenzie. 1995. An analysis of potential legal liability incurred through audit expert systems. *Intelligent Systems in Accounting, Finance, and Management* 4: 191–204.
- Whitcotton, S. M. 1996. The effects of experience and confidence on decision aid reliance: A causal model. *Behavioral Research in Accounting* 8: 194–216.
- Zimbelman, M. F. 1997. The effects of SAS No. 82 on auditors' attention to fraud risk factors and audit planning decisions. *Journal of Accounting Research* 35 (Supplement): 75–97.

Further Evidence on the Extent and Origins of JIT's Profitability Effects

Michael R. Kinney

Texas A&M University

William F. Wempe

Texas Christian University

ABSTRACT: Empirical research provides scant evidence that just-in-time (JIT) adopters outperform their non-adopting industry peers. Using a sample of 201 JIT adopters and matched non-adopters, we examine the relation between financial performance and JIT. Our sample-wide results indicate that JIT adopters improve financial performance relative to non-adopters, and that profit margin, rather than asset turnover, is the primary source of such improvement. However, results of additional analyses suggest that JIT adopters below a firm-size threshold do not improve financial performance, a finding that reconciles our study to Balakrishnan et al. (1996), which examined a JIT adopter sample that included a greater proportion of small firms.

Keywords: *just-in-time; inventory management; production technology.*

Data Availability: *All data used in this study are available from public sources.*

I. INTRODUCTION

Since the early 1980s, many U.S. firms have adopted just-in-time production management (JIT). However, empirical evidence on the association between JIT adoption and financial performance is mixed. Although Young and Selto (1991) concluded from early, site-based evidence that JIT helped adopters reduce non-value-added costs, more

For their helpful comments on earlier versions of the paper, we thank Ed Swanson, Bill Cready, Bob Strawser, Mike Wilkins, Tom Lopez, Robert Trezevant, Ramji Balakrishnan, Mahendra Gupta, Nick Dopuch, Bill Rankin, Jim Hesford, Bart Hamilton, Cecily Raiborn, and two anonymous referees, as well as workshop participants at Texas A&M University, Texas Christian University, Washington University in St. Louis, University of Nebraska, Western Michigan University, the 1999 American Accounting Association Southwest Regional Meeting, and the 1999 American Accounting Association Annual Meeting. We are grateful to the authors of Balakrishnan, Linsmeier, and Venkatachalam (1996) for providing us with their sample. We also gratefully acknowledge the research assistance of Jonathan Anderson, Jap Efendi, and Doug Lindauer.

Submitted May 1999

Accepted May 2001

recent empirical studies (e.g., Huson and Nanda 1995; Balakrishnan et al. 1996) have generally found that JIT does not necessarily lead to improved financial performance.

We re-examine the association between JIT adoption and financial performance. Our study most closely resembles Balakrishnan et al. (1996) (hereafter BLV). Using samples of 46 JIT adopters and matched non-adopters, BLV concluded that adopters improved inventory turnover relative to non-adopters. However, BLV did not find that JIT adoption, on average, is associated with improved return on assets (ROA). Instead, BLV's evidence suggests that JIT-related ROA improvement is restricted to adopters with diffuse customer bases, which outperform adopters with concentrated customer bases and non-adopting industry counterparts. BLV inferred that "captive" firms with one or more major customers may not retain JIT's financial benefits, but instead may be forced to pass along such benefits to their more powerful customers.¹

We extend BLV's study by using a larger and more recent sample to examine the association between JIT adoption and changes in financial performance. We also investigate whether JIT adopters' relative ROA changes are primarily attributable to changes in either the profit margin or asset turnover component of ROA.

We identify JIT adopters and match them with non-adopting firms, comparing changes in their respective financial performances from a three-year pre-adoption period to a three-year post-adoption period. To validate our sample selection, we first examine the relation between JIT adoption and changes in both inventory turnover and inventory-to-total assets ratios. We find that adopters enjoy an inventory turnover increase six to eight times greater than that of non-adopters, representing a 16 to 20 percent increase over pre-adoption turnover levels. JIT adopters also reduce inventory-to-total assets more substantially than do non-adopters.

In contrast to BLV, our results suggest that, on average, JIT adopters improve ROA relative to non-adopters, and that this relation does not vary with the concentration of adopters' customer bases. Moreover, our results indicate that JIT adopters improve both the profit margin and asset turnover components of ROA, relative to their non-adopting counterparts. We then compare the relative contributions of changes in profit margin and asset turnover to JIT adopters' relative ROA changes. We find that adopters' relatively improved ROAs derive largely from improved profit margins, suggesting that JIT's primary benefits stem from the elimination of non-value-adding production costs, rather than from mere reductions in total investment arising from leaner inventories. This finding regarding the source of ROA improvement is consistent with Alles et al. (1995), who suggest that the absence of buffer stocks increases the incentive to eliminate production problems, thereby eliminating the costs associated with such problems.

¹ Several other studies examine the association between JIT adoption and profitability. Anyane-Ntow (1991) examines six samples of 50 JIT adopters in six industrialized nations. The author's analysis of the 1984–1986 period indicates that Japanese JIT adopters maintain lower inventory levels compared to adopters in other nations, but do not enjoy higher profitability. Huson and Nanda (1995) examine a sample of 55 firms that adopted JIT during the 1980s and find average EPS deterioration of 25 percent for adopters, and 14 percent for matched non-adopters in the four-year post-adoption period (although in a five-equation system, the reduced form estimate of the effect of JIT adoption on EPS is positive). Callen et al. (2000) examine a sample of 60 manufacturing plants that adopted JIT between 1985 and 1989, and conclude that JIT is associated with lower costs and higher profits. Biggart (2000) examines a sample of 95 firms that adopted JIT between 1975 and 1995, and reports that changes in ROA were no different for JIT adopters and non-adopting industry counterparts. Finally, Fullerton and McWatters (2001) examine JIT's profitability effects using publicly available financial data combined with a proprietary survey of 253 firms, 95 of which were pre-1995 JIT adopters. The authors document a positive relation between profitability and JIT implementation.

Two differences between our study and BLV lead to fundamental differences in the studies' JIT adopter samples. First, as an early study examining the relation between financial performance and JIT, BLV used a sample of 46 firms that adopted JIT prior to 1990. In contrast, our sample of 201 adopters includes not only such "early" adopters, but also 78 post-1989 adopters. Second, BLV focused (in part) on customer concentration's effect on the relation between JIT and financial performance; accordingly, BLV limit their sample to firms reporting operations in fewer than four three-digit SICs (Balakrishnan et al. 1996, 189). Our primary interest is the sample-wide association between JIT and financial performance, and we examine the roles of customer concentration and other possible intervening variables in an *ex post* attempt to refine our understanding of that association. Thus, we employ no *ex ante* span-of-operations screen.

These differences in research timing and focus contribute to a significant between-study difference in sample size. Our sample of adopters is more than four times larger than BLV's, which increases the power of our tests. Moreover, BLV's span-of-operations screen, and our omission of such a screen, contributes to a substantial difference in average JIT adopter size between the two studies—our adopters are significantly larger. We find that this difference in adopter size basically accounts for the two studies' differing sample-wide ROA results. Our analyses indicate that in both studies' samples, JIT adopters below a minimum firm-size threshold achieve no relative ROA improvement, whereas firms above this threshold (i.e., medium-size to large firms) do achieve relative improvement. These findings support both Alles et al.'s (1995) contention that JIT's benefits may be greatest where there are substantial asymmetries between workers and management, as well as BLV's observation that firms with more market power may retain more of the financial benefits of JIT adoption.

The remainder of this paper is organized as follows. In the next section, we develop four hypotheses regarding associations between JIT adoption and changes in financial performance. Section III describes our sample selection procedures and provides sample descriptive statistics. Section IV explains our hypotheses tests and reports test results. Section V describes sensitivity tests and related results, and reconciles our findings with BLV's. The final section summarizes the paper and offers suggestions for future research.

II. HYPOTHESES

Profitability Hypotheses

McIlhatten (1987, 23) defines JIT as "the constant and relentless pursuit of the elimination of waste, with waste being defined as anything that does not add value to a product—inspection, queue time, and stock." Lubben (1988, 13) asserts that "the greatest misconception about JIT is that it is an inventory control system. Although structuring a system for JIT will control inventory, that is not its major function." Thus, our primary tests employ a comprehensive measure of firm performance, ROA, instead of a measure capturing only inventory management. ROA ($\text{income} \div \text{average total assets}$) is the product of profit margin ($\text{income} \div \text{sales}$) and asset turnover ($\text{sales} \div \text{average total assets}$). Although interdependent, profit margin and asset turnover reflect different determinants of a firm's success or failure. Atkinson et al. (2001, 543) describe asset turnover as a measure of productivity—the ability to generate sales with a given level of investment, and profit margin as a measure of efficiency—the ability to control costs at a given level of sales activity.

We expect JIT adoption to be associated with improved asset turnover. A JIT adopter's ability to turn its assets should increase with improved product quality, greater responsiveness to customer demand because of shorter lead times, and the greater product line variety

that is sustainable with increased manufacturing flexibility. In addition, lower inventory also reduces the denominator used to calculate total asset turnover.²

We also expect that JIT is associated with increased profit margins. JIT utilization reveals activities that add no value. These activities and their related costs are often either hidden by large inventory stocks, or are ignored because buffer stocks are a convenient "solution" to production line or other system failures. Without excess inventory to mitigate these problems, JIT adopters are more likely to develop cost-saving solutions, thereby increasing profit margin. Alles et al. (1995, 188) note:

Eliminating buffer inventories makes the production setting "transparent," exposing flaws and thus helping management and workers to eliminate problems. The incentive to eliminate these problems is greater when inventories are low, because small buffer stocks provide less insurance against problems. Management's role is to support and coach workers. By providing this support, managers get better information about problems faced by workers and the ways workers overcome these problems.

The preceding discussion suggests that JIT adopters increase ROA and both of its components, relative to the performance of non-adopting control firms:

H1: Pre- to post-adoption change in ROA is greater for JIT adopters than for matched non-adopters.

H2: Pre- to post-adoption change in asset turnover is greater for JIT adopters than for matched non-adopters.

H3: Pre- to post-adoption change in profit margin is greater for JIT adopters than for matched non-adopters.

Predominant Source of JIT-Related ROA Performance

We extend prior literature by considering whether relative changes in JIT adopters' ROAs result primarily from relative changes in efficiency (i.e., profit margin) or from relative changes in productivity (i.e., asset turnover). Because we cannot predict which effect dominates, our fourth hypothesis is nondirectional:

H4: Relative changes in profit margin and asset turnover are equally significant in explaining relative ROA changes associated with JIT adoption.

III. SAMPLE SELECTION AND DESCRIPTIVE STATISTICS

Sample Selection

Table 1 summarizes our sample selection procedures, which included searches of both public data sources and existing literature. Inclusion in the initial adopter sample required only that "JIT" or some variation thereof appear in a publicly available document (e.g., an annual report or news story) related to a given firm. The initial search identified 623 firms, of which 445 had at least five years of sales data available on Compustat. After carefully reading the 445 source documents that tentatively identified firms as JIT adopters, we concluded that 130 firms had not integrated JIT into their operations; we deleted these firms

² The effect of JIT implementation on asset turnover depends on how the firm uses any capital released via inventory reductions. For example, JIT adopters that substitute noninventory capital investment (e.g., manufacturing cells) for inventory investment may reduce or eliminate any otherwise observable improvement in total asset turnover.

TABLE 1
Sample Selection and Screening

	<i>Sample Size</i>
Searched Lexis/Nexis COMPANY file ^a	408
Searched Lexis/Nexis ALLNEWS file ^a	142
Searched NAARS ^a	9
Additional firms identified during the literature review	55
Additional firms provided by Balakrishnan et al. (1996)	9
Total potential sample firms	623
Firms without five years of data available on Compustat (1977–1995) ^b	–178
Firms for which we examined source documents (e.g., annual reports)	445
Firms eliminated because source indicated only a passing reference to JIT or a JIT adoption date outside the range of years for which Compustat data are available ^c	–130
Firms for which inventory valuation method was not available on Compustat	–4
Firms for which we could not find a match on Compustat ^d	–9
Preliminary sample	302
Firms missing data necessary to calculate total inventory turnover and ROA in each year from –3 through +3 relative to the JIT adoption year	–101
Final sample	201

^a The COMPANY file included detailed 10-Ks back to 1987 only. We searched the NAARS and ALLNEWS files to obtain pre-1987 adopters (although NAARS does not contain the Management Discussion & Analysis portion of firms' annual reports). The search string was ((just in time) or (JIT) or (pull system) or (continuous flow manufacturing) or (zero inventor!)) w/10 ((implement!) or (chang!) or (switch!) or (adopt!) or (enhanc!) or (expand!) or (extend!)).

^b Many firms identified in the ALLNEWS file and during the literature review were private firms or subsidiaries or divisions of other firms. 151 of the 178 firms that this screen eliminated had no Compustat data.

^c Passing references to JIT ranged from use of the phrase "just in time" in contexts unrelated to JIT, to discussions regarding supplying customers on a JIT basis with no indication that the firm had adopted JIT practices in its own operations. We also eliminated two firms that appeared to be unlikely candidates for implementing JIT. One was in SIC 4833 (Television Broadcast Stations); the other was in SIC 9995 (Nonoperating Establishments).

^d These firms could not be matched on inventory valuation method at the two-digit industry level.

from the sample. Of the remaining 315 firms, we deleted 13 because we could not determine an inventory valuation method (4 firms) or because we could not match (as discussed below) on inventory valuation method at the two-digit level or higher (9 firms). Our preliminary sample, therefore, includes 302 JIT adopters.

We matched non-adopting firms to the 302 JIT adopters in the preliminary sample. We matched first on inventory valuation method, and then attempted to match on net sales in the year preceding JIT adoption at the four-digit industry level. If an acceptable match was not available, then we dropped to the three-digit level and, if necessary, to the two-digit

level.³ We followed BLV in defining pre-adoption and post-adoption windows and deleted 101 firm-pairs for which either the JIT adopter or its matched control firm lacked Compustat data to calculate inventory turnover and ROA for each year from -3 to $+3$, relative to the JIT adoption year. Thus, the final sample includes 201 JIT adopters and matched counterparts.

Three features of our sample selection may affect our test results. First, firms often implement JIT incrementally, and the stage and scope of implementation can be difficult to determine from public sources. Thus, for some firms, our "JIT adoption year" is an estimate based on our reading of adopters' source documents. Second, our tests, like those of prior studies, may be subject to selection bias. If the firm characteristics that led to JIT adoption also led to superior future performance for reasons unrelated to JIT, then any performance effects associated with JIT adoption may actually be caused by such characteristics. Although sensitivity analyses reported in Section V yield no evidence that our inferences are affected by selection bias, limitations in the research design preclude us from completely ruling out such bias as an alternative explanation for our results. Third, our sample could also suffer from a disclosure bias if only successful adopters disclose implementation of JIT in their annual reports. Our sample of adopters excludes any firms that adopted JIT, quickly deemed it a failure, and elected not to discuss the failure in their adoption-year annual reports. While this possibility creates a bias toward JIT-associated profitability in our sample, it is unlikely that the number of unsuccessful JIT adopters reaching such immediate conclusions would be large enough to account for our study's results.

Descriptive Statistics

Table 2, Panel A, provides the JIT adoption year distribution for the sample of 201 JIT adopters. The weighted-average adoption year is 1988. Table 2, Panel B reports the two-digit industry distribution for the adopter and control samples. Forty-three percent of all firm-pairs are in industries 35 or 36 (Industrial Equipment and Commercial Machinery [including computers] and Electronic Equipment, respectively). As reported in Table 2, note a, 56.7 percent of the 201 firm-pairs use LIFO to value at least some portion of total inventory.

We align the annual data of each JIT adopter and its matched non-adopter in event time based on the adoption year (year 0). Like BLV, we define the pre-adoption (post-adoption) period as the three-year period preceding (following) adoption. Following BLV, we dampen the influence of extreme observations by winsorizing at the 5th and 95th percentiles of a variable's pooled distribution (six years of pre- and post-adoption data for 402 firms).⁴ For each firm, we then measure a pre- (post-) adoption variable as its median value in the three-year pre- (post-) adoption period.

Table 2, Panel C provides means and medians of ten pre-adoption attributes for the JIT and control samples. The last four columns of Panel C are mean and median paired differences between subsamples, standard deviations of paired differences, and significance

³ We determined "acceptable" firm-size matches by judgment rather than by a predetermined criterion. We did not use a predetermined criterion because (1) the firm size/industry trade-off is difficult to quantify; (2) it is not clear that any predetermined firm size screening criterion (e.g., the ratio of adopter size to non-adopter size) would be equally appropriate for large and small adopters; and (3) controlling for firm size differences in statistical tests is straightforward. When we elected to match at a given industry level, we selected the control firm with net sales (in year -1) closest to the net sales of the adopting firm. The matching procedure yielded approximately 69 percent four-digit industry matches, 14 percent three-digit matches, and 17 percent two-digit matches.

⁴ We do not winsorize variables that are simply size measures (e.g., total assets).

TABLE 2
Descriptive Statistics for 201 JIT Adopters and Matched Control Firms^a

Panel A: Distribution of JIT Adoption Years

<u>Year</u>	<u>Number of firms</u>	<u>%</u>
1982	3	1.5
1983	6	3.0
1984	12	6.0
1985	15	7.5
1986	17	8.5
1987	18	9.0
1988	27	13.4
1989	25	12.4
1990	26	12.9
1991	22	10.9
1992	16	8.0
1993	<u>14</u>	<u>7.0</u>
	201	100.0

Panel B: Distribution of Two-Digit Industry Classifications

<u>Two-Digit Industry Code</u>	<u>Industry Description</u>	<u>Number of Firms</u>	<u>%</u>
20	Food	1	0.5
22	Textile mill products	2	1.0
23	Apparel	1	0.5
24	Lumber	1	0.5
25	Furniture	8	4.0
26	Paper	2	1.0
27	Printing, publishing	4	2.0
28	Chemicals	5	2.5
30	Rubber and plastics	5	2.5
31	Leather	2	1.0
32	Glass, pottery	1	0.5
33	Primary metals	12	6.0
34	Fabricated metals	11	5.5
35	Industrial equipment	46	22.9
36	Electronic equipment	40	19.9

(Continued on next page)

TABLE 2 (Continued)

<i>Two-Digit Industry Code</i>	<i>Industry Description</i>	<i>Number of Firms</i>	<i>%</i>
37	Motor vehicles	16	8.0
38	Instrumentation	23	11.4
39	Other manufacturing	6	3.0
50	Wholesale durables	4	2.0
53	Department stores	3	1.5
54	Grocery stores	2	1.0
56	Clothing stores	1	0.5
57	Furniture stores	1	0.5
59	Miscellaneous retail	2	1.0
73	Packaged software	2	1.0
		201	100.0

Panel C: Financial Attributes of the JIT Adopter and Control Samples

<i>Firm Attribute^b</i>	<i>JIT Sample</i>		<i>Control Sample</i>		<i>Paired Difference</i>			
	<i>Mean</i>	<i>Median</i>	<i>Mean</i>	<i>Median</i>	<i>Mean</i>	<i>Median</i>	<i>Std. Dev.</i>	<i>p-value^c</i>
Total assets	1,736.1	303.3	988.0	169.0	748.1	66.2	3,876.5	0.001
Net sales	2,365.0	397.5	1,185.8	234.3	1,179.2	79.0	5,803.9	0.001
Total inventory	348.4	71.2	196.9	43.0	151.6	16.4	679.5	0.001
Gross margin	0.349	0.332	0.333	0.312	0.016	0.014	0.136	0.108
Operating ROA	0.107	0.110	0.107	0.104	0.000	0.004	0.086	0.978
ROA	0.056	0.056	0.057	0.057	-0.001	0.001	0.063	0.861
Inventory ÷ assets	0.251	0.236	0.257	0.249	-0.006	0.000	0.120	0.472
Inventory turnover	4.041	3.437	4.200	3.601	-0.159	0.000	2.398	0.592
Debt ÷ assets	0.486	0.488	0.492	0.493	-0.006	-0.029	0.221	0.504
Fixed-cost ratio	0.062	0.053	0.052	0.044	0.009	0.004	0.038	0.002

^a The adopter and control samples match on inventory valuation method. We coded firms as LIFO users if any digit in Compustat item #59 indicated LIFO usage; otherwise, we coded firms as FIFO users. 114 (56.7 percent) of all firm-pairs are LIFO users.

^b Firm attribute measures are median values in the three-year period preceding the adoption year. Total assets, net sales, and total inventory are stated in \$ millions. Operating ROA is calculated with operating income after depreciation in the numerator, whereas ROA is calculated with income before extraordinary and special items in the numerator (as in BLV). Both measures use average total assets in the denominator. Inventory ÷ assets = total inventory ÷ total assets. Inventory turnover = cost of sales ÷ average total inventory. Debt ÷ assets = total debt ÷ total assets. Fixed-cost ratio = depreciation ÷ cost of sales.

^c p-values are the significance levels from two-tailed Wilcoxon signed rank tests.

levels from Wilcoxon signed rank tests.⁵ Paired differences in total assets, net sales, and total inventory indicate that JIT adopters are, on average, larger than their non-adopting counterparts. Aside from this size difference, the only other significant difference in Panel C is for fixed cost (depreciation ÷ cost of sales) ratios, with JIT adopters having the larger ratios.

Sample Validation

To validate our sample selection procedures, we examine whether our JIT sample firms differ from non-adopting matched firms in ways consistent with JIT adoption. In Figure 1, we plot median total inventory turnover (cost of sales ÷ average total inventory) for both samples in years -3 through +3. The graph indicates that adopters increased total inventory turnover (both absolutely and in comparison to non-adopters), and also suggests that adoption years are reasonably well identified. Although inventory turnover mildly increases in year -1, the first substantial increase occurs in year 0, followed by sustained increases in subsequent years.

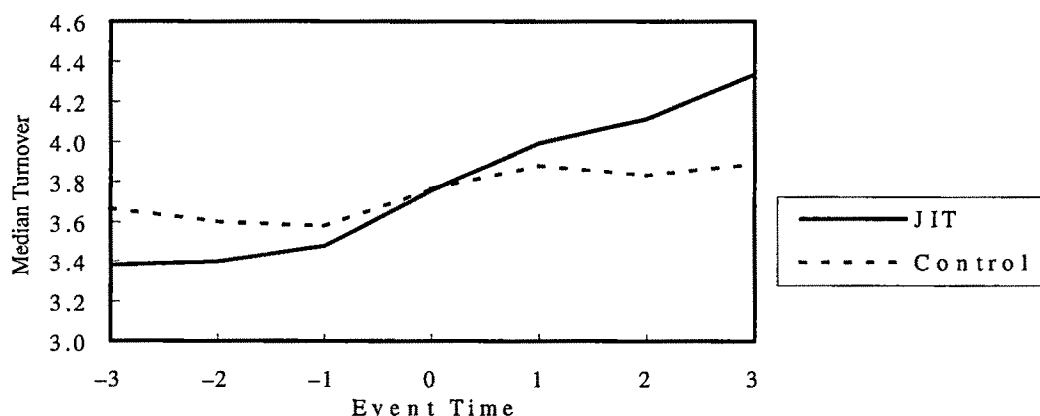
To validate our sample selection procedures more formally, we conduct Wilcoxon signed rank tests of the paired differences in changes in total inventory turnover and investment:

$$\text{DIFAINVTURN}_j = \Delta \text{INVTURN}_j - \Delta \text{INVTURN}_C \quad (1)$$

$$\text{DIFAINVASSET}_j = \Delta \text{INVASSET}_j - \Delta \text{INVASSET}_C \quad (2)$$

⁵ Later tables reporting test results present only medians and the results of Wilcoxon signed rank tests. However, means and related t-tests produced identical inferences.

FIGURE 1
Changes in Total Inventory Turnover for 201 JIT and Control Firms^a



^a Total inventory turnover is defined as cost of sales divided by average total inventory. Event time is the year relative to the JIT adoption year.

where:

J = JIT adopter;

C = matched control firm;

$\Delta \text{INVTURN}$ = post-adoption total inventory turnover minus pre-adoption total inventory turnover; and

$\Delta \text{INVASSET}$ = post-adoption total inventory $\div$ total assets minus pre-adoption total inventory $\div$ total assets.

Table 3 results confirm that JIT adopters achieve greater increases in total inventory turnover and greater reductions in inventory investment. For the JIT sample, both changes are significant at $p \leq 0.001$ in one-tailed signed rank tests. Tests of paired differences between the JIT and control samples indicate that pre-adoption differences in inventory turnover and investment are not significant. However, in the post-adoption period, JIT adopters have higher inventory turnover and lower inventory investment ($p = 0.012$ and 0.002 , respectively). Finally, tests of paired differences in changes indicate that JIT adopters increased inventory turnover and reduced inventory investment, compared to their matched counterparts ($p \leq 0.001$ in both cases).⁶

⁶ We also tested raw material, work-in-process, and finished goods inventory components. In the JIT sample, turnover increased for all three components ($p \leq 0.007$ for each component). Moreover, raw materials and work-in-process inventory turnover increased more for adopters than for control firms ($p = 0.030$ and 0.046 , respectively).

TABLE 3
Tests of Changes in Total Inventory Turnover and Investment^a

	<i>JIT Median</i>	<i>Control Median</i>	<i>Median Paired Difference</i>	<i>p-value^b</i>
Total Inventory Turnover				
Pre-adoption	3.437	3.601	0.000	0.592
Post-adoption	4.115	3.891	0.482	0.012
Change	0.407	0.073	0.401	0.001
p-value, intra-sample change ^b	0.001	0.085		
Total Inventory $\div$ Total Assets				
Pre-adoption	0.236	0.249	0.000	0.472
Post-adoption	0.190	0.217	-0.022	0.002
Change	-0.032	-0.007	-0.021	0.001
p-value, intra-sample change ^b	0.001	0.001		

^a Total inventory turnover is cost of sales divided by average total inventory. For a given firm, pre-adoption total inventory turnover and total inventory $\div$ total assets are median values of the respective measures in the three years preceding the JIT adoption year. Post-adoption total inventory turnover and total inventory $\div$ total assets are median values of the respective measure in the three years following the JIT adoption year.

^b For the adopter sample, p-values for intra-sample changes are from one-tailed signed rank tests. For the control sample, p-values for intra-sample changes are from two-tailed signed rank tests. p-values for pre-adoption differences are from two-tailed signed rank tests. p-values for post-adoption differences and differences in changes in performance are from one-tailed signed rank tests. Tests of paired differences in changes in inventory turnover and inventory $\div$ assets are tests of $\Delta \text{INVTURN}$ and $\Delta \text{INVASSET}$, as presented in Equations (1) and (2), respectively. We conducted all analyses using 201 firm-pairs.

IV. H1–H4 TESTS AND TEST RESULTS

Tests of the ROA, Asset Turnover, and Profit Margin Hypotheses (H1–H3)

Our tests of changes in ROA, asset turnover, and profit margin are analogous to the sample validation test:

$$\text{DIF}\Delta\text{ROA}_J = \Delta\text{ROA}_J - \Delta\text{ROA}_C \quad (3)$$

$$\text{DIF}\Delta\text{ASSETTURN}_J = \Delta\text{ASSETTURN}_J - \Delta\text{ASSETTURN}_C \quad (4)$$

$$\text{DIF}\Delta\text{MARGIN}_J = \Delta\text{MARGIN}_J - \Delta\text{MARGIN}_C \quad (5)$$

where:

J and C = JIT and control firms, respectively;

ΔROA = post-adoption ROA minus pre-adoption ROA, where ROA is income before extraordinary and special items $\div$ average total assets;⁷

$\Delta\text{ASSETTURN}$ = post-adoption asset turnover minus pre-adoption asset turnover, where asset turnover is net sales $\div$ average total assets; and

ΔMARGIN = post-adoption profit margin minus pre-adoption profit margin, where profit margin is income before extraordinary and special items $\div$ net sales.

We test H1–H3 with one-tailed signed rank tests of $\text{DIF}\Delta\text{ROA}$, $\text{DIF}\Delta\text{ASSETTURN}$, and $\text{DIF}\Delta\text{MARGIN}$, respectively.

ROA (H1) Results

Table 4 indicates that JIT adopters' ROAs declined an average of 3.6 percent $([0.054 - 0.056] \div 0.056)$ from the pre- to post-adoption period, which is significant in a two-tailed signed rank test ($p = 0.055$). However, non-adopters' ROAs decline much more—

⁷ Operating income yielded identical inferences.

TABLE 4
Tests of Return on Assets Changes for JIT and Control Samples (H1)

	<i>JIT Median</i>	<i>Control Median</i>	<i>Median Paired Difference</i>	<i>p-value^b</i>
Return on Assets ^a				
Pre-adoption	0.056	0.057	0.001	0.861
Post-adoption	0.054	0.045	0.007	0.010
Change	−0.005	−0.015	0.010	0.010
p-value, intra-sample change ^b	0.055	0.001		

^a Pre-adoption ROA is median ROA in the three years preceding the JIT adoption year. Post-adoption ROA is median ROA in the three years following the JIT adoption year.

^b For intra-sample tests, p-values are from two-tailed signed rank tests. For paired differences, the p-value in the test of pre-adoption ROA is from a two-tailed signed rank test, and p-values for tests of post-adoption ROA and changes in ROA are from one-tailed tests. The paired difference in change in ROA is $\text{DIF}\Delta\text{ROA}$, as presented in Equation (3).

by 21.1 percent ($p = 0.001$).⁸ The last two columns of Table 4 report tests of paired differences. In the pre-adoption period, there was no significant difference between adopters and control firms ($p = 0.861$), but adopters' ROAs were higher than control firms' ROAs in the post-adoption period ($p = 0.010$). Most importantly, the results support H1: JIT adopters' ROAs increase relative to control firms' ROAs ($p = 0.010$).⁹

We also tested paired ROA changes separately for years 0 through +3. For these years, median paired differences in ROA changes were 0.006, 0.007, 0.007, and 0.003, respectively. Thus, JIT adopters' relative ROA improvement deteriorates by the end of the post-adoption period. We conducted additional tests on years +4, +5, and +6 using 128 firm-pairs for which data were available through year +6. For year +4, the paired differences in ROA changes were significant at the 0.078 level; for years +5 and +6, the paired differences were insignificant. This apparent dissipation of the association between JIT and financial performance is consistent with BLV's (1996, 187) argument that, in a competitive market, firms will mimic early JIT adopters if JIT adoption improves performance. In the long run, any gains realized by early adopters will be competed away. Thus, control firms may have ultimately adopted JIT, but made no related disclosures because JIT adoption was no longer newsworthy at the time of their late adoptions.¹⁰ Consistent with this scenario, paired differences in changes in inventory turnover are also no longer significant by year +6, although inventory turnover for both samples improved greatly.

Asset Turnover (H2) and Profit Margin (H3) Results

Results from tests of H2 and H3 appear in Table 5. JIT adopters' pre- to post-adoption asset turnover changes were statistically insignificant, whereas asset turnover declined significantly for non-adopters ($p = 0.013$). The tests of pre- to post-adoption changes in profit margin indicate that declines occurred in both samples ($p = 0.114$ and 0.001 for adopters and non-adopters, respectively). Results of paired tests indicate there were no pre-adoption differences in asset turnover and profit margin between samples. Post-adoption results reveal that profit margin is significantly higher for adopters ($p = 0.010$). Most important, our H2 prediction that relative asset turnover would improve for JIT adopters is supported ($p = 0.033$), and our H3 prediction that JIT adopters would enjoy a higher relative change in profit margin is also supported ($p = 0.014$).¹¹

Tests of the Dominant Source of Post-Adoption ROA Performance (H4)

Tests of H4 examine the relative effects of profit margin and asset turnover changes on ROA changes. We first construct two variables that capture the ROA effect of relative changes in profit margin and asset turnover:

$$\text{MARGIN\_EFFECT}_j = \text{DIF}\Delta\text{MARGIN}_j \times \text{PRE\_ASSETTURN}_j \quad (6)$$

$$\text{TURNOVER\_EFFECT}_j = \text{DIF}\Delta\text{ASSETTURN}_j \times \text{PRE\_MARGIN}_j \quad (7)$$

⁸ As in BLV, macroeconomic factors likely account for within-sample ROA declines. In BLV, the adopter (control) sample experienced a 14.3 percent (36.2 percent) pre- to post-adoption ROA decline.

⁹ We examined the extent to which firms with pre-adoption losses affect our results. We first deleted firm-pairs in which the adopters reported pre-adoption losses ($n = 25$). Hypothesis 1 was supported at the 0.069 level. We then deleted (in addition to these 25 firm-pairs) firm-pairs in which the non-adopters reported pre-adoption losses ($n = 14$). In the remaining sample of 162 firm-pairs, H1 was supported at the 0.008 level.

¹⁰ Additional electronic searches indicated that through year +6, only six firms in our control sample ultimately reported implementation of JIT practices.

¹¹ After deleting firm-pairs for which adopters report pre-adoption losses ($n = 25$), the asset turnover and profit margin hypotheses are supported at the 0.052 and 0.125 levels, respectively. When we also delete firm-pairs with non-adopters reporting such losses ($n = 14$), the asset turnover and profit margin hypotheses are supported at the 0.012 and 0.015 levels, respectively.

TABLE 5
Tests of Asset Turnover (H2) and Profit Margin (H3) Changes for JIT and Control Samples

	<i>JIT Median</i>	<i>Control Median</i>	<i>Median Paired Difference</i>	<i>p-value^b</i>
Asset Turnover ^a				
Pre-adoption	1.360	1.365	0.005	0.867
Post-adoption	1.348	1.332	0.000	0.215
Change	0.000	-0.018	0.035	0.033
p-value, intra-sample change ^b	0.845	0.013		
Profit Margin ^a				
Pre-adoption	0.040	0.039	0.001	0.825
Post-adoption	0.038	0.030	0.004	0.010
Change	-0.001	-0.005	0.005	0.014
p-value, intra-sample change ^b	0.114	0.001		

^a Pre-adoption asset turnover and profit margin are the median values of the respective measures in the three years preceding the JIT adoption year. Post-adoption asset turnover and profit margin are the median values of the respective measures in the three years following the JIT adoption year.

^b For intra-sample tests, p-values are from two-tailed signed rank tests. For paired differences, p-values in tests of pre-adoption performance are from two-tailed signed rank tests, and p-values in tests of post-adoption performance and changes in performance are from one-tailed tests. Paired differences in changes in asset turnover and profit margin are DIFΔASSETTURN and DIFΔMARGIN, as presented in Equations (4) and (5), respectively.

where DIFΔMARGIN and DIFΔASSETTURN are as previously defined, PRE_ASSETTURN is adopter *j*'s pre-adoption asset turnover, and PRE_MARGIN is adopter *j*'s pre-adoption profit margin.

With a signed rank test of MARGIN_EFFECT, we examine the ROA effect of relative profit margin changes, when JIT adopters' asset turnovers are held constant at pre-adoption levels. We conduct a similar test for asset turnover. Our test of H4 examines within-adopter differences in MARGIN_EFFECT and TURNOVER_EFFECT:

$$\text{DIF\_EFFECT}_j = \text{MARGIN\_EFFECT}_j - \text{TURNOVER\_EFFECT}_j. \quad (8)$$

Significantly positive DIF_EFFECT would suggest that relative profit margin changes, compared to relative asset turnover changes, more positively influence JIT adopters' post-adoption relative ROA changes. Significantly negative DIF_EFFECT would suggest the converse is true. We test DIF_EFFECT with a two-tailed signed rank test.

Source of ROA Performance (H4) Results

Table 6 reports results of H4 tests conducted on full (*n* = 201), once-reduced (*n* = 176), and twice-reduced (*n* = 162) samples. The once-reduced sample excludes the 25 JIT adopters reporting pre-adoption losses. For these 25 firms, TURNOVER_EFFECT is negative for adopters with relative increases in asset turnover, and positive for adopters with relative decreases in asset turnover. Although these effects are consistent with the DuPont formula (e.g., an unprofitable firm that increases its asset turnover will decrease ROA), the presence of these firms in the full sample may understate the effect of asset turnover on JIT adopters' relative ROA changes. In addition, excluding these 25 firms eliminates the effect of mean reversion on the part of those JIT adopters that were least successful during the pre-adoption period (although, arguably, such firms also have the

TABLE 6
ROA Effects of JIT Adopters' Relative Profit Margin and Asset Turnover Changes (H4)^a

	<i>Full Sample^b</i> <i>n = 201</i>		<i>Once-Reduced Sample^b</i> <i>n = 176</i>		<i>Twice-Reduced Sample^b</i> <i>n = 162</i>	
	<i>Median</i>	<i>p-value^c</i>	<i>Median</i>	<i>p-value^c</i>	<i>Median</i>	<i>p-value^c</i>
MARGIN_EFFECT	0.006	0.014	0.004	0.113	0.009	0.015
TURNOVER_EFFECT	0.001	0.322	0.001	0.177	0.002	0.063
DIF_EFFECT	0.007	0.032	0.002	0.259	0.005	0.052

^a We calculate the ROA effect of relative changes in profit margin and asset turnover as follows:

MARGIN_EFFECT = $\text{DIF}\Delta\text{MARGIN} \times \text{JIT adopter's pre-adoption asset turnover}$, where $\text{DIF}\Delta\text{MARGIN}$ is the paired difference in the change in profit margin (see Equation [6]).

TURNOVER_EFFECT = $\text{DIF}\Delta\text{ASSETTURN} \times \text{JIT adopter's pre-adoption profit margin}$, where $\text{DIF}\Delta\text{ASSETTURN}$ is the paired difference in the change in asset turnover (see Equation [7]).

DIF_EFFECT = MARGIN_EFFECT – TURNOVER_EFFECT is tested to assess the relative importance of profit margin and asset turnover performance in JIT adopters' relative ROA changes (see Equation [8]).

^b The full sample is the 201 JIT adopters used in primary tests. The once-reduced sample is identical, except that JIT adopters with negative pre-adoption profit margins (25 firms) are removed. The twice-reduced sample is identical to the once-reduced example, except that JIT adopters paired with control firms with pre-adoption losses (14 firms) are also removed.

^c p-values are from two-tailed signed rank tests for DIF_EFFECT. Other p-values are from one-tailed signed rank tests.

most to gain from JIT-induced improvement in cost control). The twice-reduced sample excludes these 25 firms (thus purging the somewhat perverse TURNOVER_EFFECT) and the 14 JIT adopters paired with non-adopters reporting pre-adoption losses (thereby applying any concern regarding mean reversion uniformly to the adopter and control samples).

In the tests of the full sample, median MARGIN_EFFECT is significantly positive ($p = 0.014$), but median TURNOVER_EFFECT is insignificant ($p = 0.322$). Most important, median DIF_EFFECT is positive (0.007) and significant at the 0.032 level in a two-tailed test, indicating that profit margin changes, compared to asset turnover changes, contribute more to JIT adopters' relative ROA improvement. As might be expected, the tests conducted on the once-reduced sample do not fully support the primacy of profit margin in JIT adopters' relative ROA improvement—median DIF_EFFECT, though positive, is not significant ($p = 0.259$). Tests of the twice-reduced sample suggest that relative improvement in profit margin and asset turnover both contribute to relative ROA improvement ($p = 0.015$ and 0.063, respectively), but that the ROA effect of profit margin changes exceeds that of asset turnover changes ($p = 0.052$ in the two-tailed test). Overall, the results in Table 6 suggest that the positive association between JIT and ROA performance results primarily from JIT adopters' relatively improved profit margins, consistent with Alles et al.'s (1995) contention that JIT promotes transparent operations and cost-saving improvements in production processes.

V. SENSITIVITY TESTS AND RECONCILIATION WITH BLV

Self-Selection Sensitivity Test

Our primary tests ignore self-selection inherent in the JIT adoption decision. Therefore, if the underlying firm characteristics that prompt JIT adoption are also associated with

superior future performance (without regard to JIT's performance effects), then our tests of paired differences in changes in ROA, asset turnover, and profit margin may overstate JIT's performance effects. To evaluate the effects of endogeneity on our H1 (ROA) inferences, we conduct two-stage self-selection analysis similar to that described by Maddala (1977, 1983, 1991) and used by Hogan (1997) in a study of auditor selection. The first stage in the analysis requires estimation of a probit model of JIT choice.¹² We estimated the following model:

$$\begin{aligned} \text{ADOPT}_j = & \beta_0 + \beta_1 \text{PRE\_SALES}_j + \beta_2 \text{PRE\_INVTURN}_j \\ & + \beta_3 \text{VAR\_DEMAND}_j + \beta_4 \text{INNOVATE}_j + u_j \end{aligned} \quad (9)$$

where:

ADOPT = 1 if firm *j* is a JIT adopter, and 0 otherwise;
 PRE_SALES = median sales in the pre-adoption period;
 PRE_INVTURN = median total inventory turnover in the pre-adoption period;
 VAR_DEMAND = variability of demand for the firm's output; and
 INNOVATE = firm innovativeness, proxied by median R&D ÷ Sales in the pre-adoption period.

We include PRE_SALES because average firm size differs between samples, and because firms with greater resources may be more inclined to adopt new technologies. We include PRE_INVTURN because managers of firms with low inventory turnover may anticipate the greatest benefits from JIT adoption, as BLV suggest. We include VAR_DEMAND because JIT may be less beneficial to firms for which product demand (and thus production and input demand) is more variable. To estimate VAR_DEMAND, we regress a firm's cost of sales for the 24 quarters preceding JIT adoption on a quarterly time-trend variable.¹³ VAR_DEMAND is the standard deviation of the regression residuals, scaled by mean cost of sales over the estimation period. We include INNOVATE to reflect the possibility that innovative firms, proxied by R&D spending scaled by sales, may be more likely to adopt new technologies. Finally, we estimate the model with samples matched on industry, inventory method, and size.

The upper portion of Table 7 reports results from estimating the first-stage model. Coefficient estimates all have the expected signs, and the coefficients of PRE_SALES and INNOVATE are significant ($p = 0.050$ and 0.010 , respectively). In addition, the model is significant at the 0.001 level.

In the second stage of the selectivity analysis, we estimated, separately for the two samples, the following model (firm subscripts are suppressed):

¹² Maddala (1977, 351–366; 1983, 257–290) explains the theoretical basis for the procedure. Our study uses mean $[\Delta \text{ROA}_j - \Delta \text{ROA}_C]$, the mean paired difference in change in ROA from our sample, as an estimate of what we are really interested in, $\mu_j - \mu_C$, which is the mean ROA if *all* firms adopt JIT, minus the mean ROA if *no* firms adopt JIT. Suppose that JIT adopters are simply superior to non-adopters, even without adopting JIT. If so, mean ΔROA_j is an upward-biased estimate of μ_j (i.e., if non-adopters had adopted JIT and were in our adopter sample, mean ΔROA_j would be smaller), whereas mean ΔROA_C is a downward-biased estimate of μ_C (i.e., if adopters had not adopted JIT, but were instead in our control sample, mean ΔROA_C would be larger). As a result, mean $[\Delta \text{ROA}_j - \Delta \text{ROA}_C]$ would be an upward-biased estimate of $\mu_j - \mu_C$. We use Maddala's two-stage procedure to discern whether selection bias is influencing our inferences in this manner.

¹³ This data requirement reduced the number of firm-pairs in this analysis to 194.

TABLE 7
Results of Two-Stage Self-Selection Analysis

Stage 1: JIT Choice Model (Equation 9)^a

$$ADOPT = \beta_0 + \beta_1 PRE\_SALES + \beta_2 PRE\_INVTURN + \beta_3 VAR\_DEMAND + \beta_4 INNOVATE + u$$

	<u>INTERCEPT</u>	<u>PRE_SALES</u>	<u>PRE_INVTURN</u>	<u>VAR_DEMAND</u>	<u>INNOVATE</u>
Pred Sign	(?)	(+)	(-)	{-}	(+)
Est. Coef.	-0.070	0.0001	-0.012	-0.451	4.74
p-value ^b	0.740	0.050	0.734	0.338	0.010

Model Significance: 0.001
n: 388 (194 firm-pairs)

Stage 2: Model Incorporating Selectivity Variables (Equation 10)^a

$$\Delta ROA = \alpha_0 + \alpha_1 PRE\_SALES + \alpha_2 PRE\_INVTURN + \alpha_3 VAR\_DEMAND + \alpha_4 INNOVATE + \alpha_5 M_{J(C)} + \varepsilon$$

	<u>INTERCEPT</u>	<u>PRE_SALES</u>	<u>PRE_INVTURN</u>	<u>VAR_DEMAND</u>	<u>INNOVATE</u>	<u>M_{J(C)}</u>
Adopters:						
Est. Coef.	-0.038	0.000002	-0.002	0.020	0.117	-0.039
t-statistic ^b	-0.86	2.40	-1.04	0.57	0.68	-0.73
Model R ² : 0.027						
Control:						
Est. Coef.	-0.123	-0.00001	0.0001	-0.013	-0.750	0.173
t-statistic ^b	-0.24	-0.21	0.01	-0.07	-0.35	0.25
Model R ² : 0.039						

^a We estimated the first-stage model with probit. ADOPT equals 1 if the firm is a JIT adopter, 0 otherwise. PRE_SALES (PRE_INVTURN) is sales (total inventory turnover) in the pre-adoption period. VAR_DEMAND is variability of demand over the 24 quarters preceding JIT adoption. INNOVATE is firm innovativeness, which we proxy by R&D ÷ Sales in the pre-adoption period. M_J and M_C are selectivity variables. For adopters, M_J is $-f(\beta'Z) \div F(\beta'Z)$; for control firms, M_C is $f(\beta'Z) \div (1 - F(\beta'Z))$, where $f(\bullet)$ and $F(\bullet)$ are the density and distribution functions, respectively, of the standard normal distribution, and $\beta'Z$ is the prediction from the first-stage probit model.

^b p-values in the first stage are from Chi-square tests. In testing coefficient estimates in the second stage, we used White's (1980) heteroskedasticity-consistent estimator of the variance-covariance matrix.

$$\Delta ROA = \alpha_0 + \alpha_1 PRE\_SALES + \alpha_2 PRE\_INVTURN + \alpha_3 VAR\_DEMAND + \alpha_4 INNOVATE + \alpha_5 M_{J(C)} + \varepsilon \quad (10)$$

where ΔROA , PRE_SALES, PRE_INVTURN, VAR_DEMAND and INNOVATE are as previously defined, and M_J and M_C are selectivity variables. In the adopter equation, M_J is $-f(\beta'Z) \div F(\beta'Z)$; for non-adopters, M_C is $f(\beta'Z) \div (1 - F(\beta'Z))$, where $f(\bullet)$ and $F(\bullet)$ are

the density and distribution functions, respectively, of the standard normal distribution, and $\beta'Z$ is the prediction from the first-stage probit model.

We assess the role of selection bias by examining the estimates of α_5 in the two equations. If $\alpha_5 < 0$ in both equations, then our primary results overstate JIT's ROA effect because ΔROA is overstated for the adopter sample and understated for the control sample. However, the coefficients of M_I and M_C , reported in the lower portion of Table 7, are negative and positive, respectively, and not significant at conventional levels. Thus, the analysis provides no evidence that self-selection bias affects our inferences. Likewise, we found no evidence of such bias in analogous H2 and H3 analyses using $\Delta ASSETTURN$ and $\Delta MARGIN$ as dependent variables in the second stage.¹⁴

Reconciliation with BLV

In contrast to BLV's findings, the results in Table 4 suggest that JIT adoption is associated with superior ROA performance, and sensitivity analyses yield no evidence that this finding is attributable to between-sample differences in firm size, cost structure, or LIFO liquidation income (see footnote 14), or to self-selection bias. In an effort to understand the between-study difference in sample-wide results, we conducted the following analyses.

Differences in Research Methods and Statistical Power

Our method, analysis of paired differences, differs from BLV's ANOVA. To examine the effect of this difference, we first ran *our* test on *BLV's sample*. Like BLV, we found no relative ROA improvement for JIT adopters. Next, we used *BLV's method* on *our* sample and observed a significant relative increase in ROA for JIT adopters. Thus, the difference in results is not method-driven, but instead is attributable to either (1) the greater statistical power of our test (i.e., our much larger sample), or (2) a fundamental difference in samples between the studies.

We ran an informal bootstrap analysis on our sample to evaluate the effect of the between-study difference in statistical power. We drew 100 random samples of size 46 (BLV's sample size) from our sample of 201 firm-pairs, and conducted our test of $DIF\Delta ROA$ on each sample. The median significance level in these tests was 0.105 (BLV's significance level in their analogous ROA test was 0.26), and 30 of 100 samples had p-values less than 0.05. Because we continue to find a reasonably significant association between JIT and ROA using these 46 firm-pair subsets of our full sample, we conclude that fundamental differences between the two studies' samples, rather than statistical power, primarily account for the studies' differential inferences.

¹⁴ In additional sensitivity analyses related to the ROA results reported in Table 4, we regressed $DIF\Delta ROA$ on paired differences in pre-adoption sales and paired differences in pre-adoption fixed-cost ratios. The coefficients on the independent variables had p-values of 0.656 and 0.086, respectively. The intercept of the equation was 0.011 ($p = 0.015$), indicating that the relative ROA increase for JIT adopters is still significant after controlling for the two significant between-sample differences identified in Table 2, Panel C. We also screened from the sample any firm-pair in which $[\text{adopter sales} - \text{non-adopter sales}] \div \text{non-adopter sales}$ in the pre-adoption period exceeded 2.5. For the remaining sample of 148 firm-pairs, the average paired difference in pre-adoption sales was not significant, and the test of $DIF\Delta ROA$ was significant at $p = 0.020$. We also examined financial statement footnotes and found that in the adoption year, 29 (17) JIT adopters (non-adopters) reported LIFO liquidation income. We re-ran the test of $DIF\Delta ROA$ after reducing firms' reported earnings by the effect of such income. Our inferences remained unchanged. We conducted the following check of our inferences' validity. We regressed $DIF\Delta ROA$ on $DIF\Delta INVTURN$. The coefficient on $DIF\Delta INVTURN$ was positive and significant at the 0.001 level (tests using, in separate equations, paired differences in changes in raw materials, work-in-process, and finished goods turnover as independent variables yielded coefficients with p-values of 0.376, 0.024, and 0.035, respectively). Finally, we note that our H2–H4 inferences are also robust to sensitivity analyses similar to those described here for ROA. The results did suggest, however, that relative asset turnover improvement was significantly greater for adopters with relatively large fixed-cost ratios.

Tests Examining Customer Concentration and Span-of-Operations Effects

BLV report significant relative ROA improvement among “free” JIT adopters not disclosing a major customer under SFAS No. 14, but no relative improvement among “captive” adopters disclosing the existence of a major customer. To enhance their customer concentration tests, BLV required that adopter and control firms match on customer concentration, and that firms in their samples report operations spanning fewer than four three-digit SICs. *Ex ante*, we elected not to impose these restrictions, thereby avoiding any resulting loss of generalizability and reduction in sample size. However, our sample includes enough firm-pairs meeting the BLV restrictions to determine whether customer concentration or span-of-operations effects likely account for differences between our results and BLV’s.

We conducted an ROA test on 114 firm-pairs that match on customer concentration, 23 captive and 91 free, and on a “quasi-matched” sample that includes these 114 firm-pairs plus 35 additional firm-pairs in which captive adopters are matched with free non-adopters. Under BLV’s hypothesis, we expect *at least* as much downstream sharing of JIT-induced cost savings in the latter 35 firm-pairs as in captive adopter/captive non-adopter firm-pairs. Thus, the test conducted on the quasi-matched sample is biased toward finding BLV’s result. We regress ΔROA on CAPTIVE (coded 1 if the JIT adopter in a firm-pair is captive, 0 otherwise), and two control variables: paired differences in pre-adoption sales ($\Delta\text{PRE\_SALES}$), and paired differences in pre-adoption fixed-cost ratios ($\Delta\text{PRE\_FCRATIO}$). BLV concluded that captive JIT firms did not enjoy relative improvement in ROA, whereas free adopters did. If this result holds in our sample, the regression equation’s intercept will be significantly positive, and CAPTIVE will have a significantly negative coefficient.

The results reported in Table 8 reveal that CAPTIVE is insignificant in both regressions, suggesting that customer concentration plays no role in JIT adopters’ relative ROA changes. Moreover, the significantly positive intercepts in the equations, along with the (insignificantly) positive coefficients on CAPTIVE, indicate that paired differences in ROA changes are significant for both free and captive JIT adopters, after controlling for between-sample differences in firm size and cost structure.¹⁵

To analyze the sensitivity of our results to a span-of-operations screen, we tested ΔROA using only those firm-pairs in which both firms operated in fewer than N three-digit SIC classes, where N , in sequential tests, was set equal to 3, 4, 5, and 6. With $N = 3$ ($N = 4$), the sample size was 60 (89) firm-pairs, and the p -value of the ROA test was 0.049 (0.108); with $N = 5$ (107 firm-pairs), the p -value was 0.037; with $N = 6$ (124 firm-pairs), the p -value was 0.026. Thus, our ROA results are not sensitive to BLV’s span-of-operations screen.¹⁶

Tests Examining Adopter Size Effects

A careful comparison of JIT adopter profiles in the BLV and present studies reveals two primary differences. First, our sample firms are much larger, on average, than BLV’s.

¹⁵ In light of the insignificance of CAPTIVE, we re-estimated the model with CAPTIVE omitted. Near the bottom of Table 8, we report intercepts (and p -values) for the reduced model. Like the ROA results reported in Table 8, additional analyses provided no evidence that customer concentration played a role in JIT adopters’ relative asset turnover and profit margin changes.

¹⁶ For each of these four reduced samples, the p -value in the test of ΔMARGIN is less than 0.05. However, p -values in the four tests of $\Delta\text{ASSETTURN}$ range from 0.113 to 0.349. We also note that our H1–H4 inferences are unchanged when we restrict our sample to firms in two-digit industry codes represented in BLV’s sample.

TABLE 8
Regression Results: Model Examining the Effect of Customer Concentration
(Free or Captive Status) on Relative ROA Changes

$$\text{Model: } \text{DIF}\Delta\text{ROA} = \beta_0 + \beta_1\text{CAPTIVE} + \beta_2\text{DIF\_PRE\_SALES} + \beta_3\text{DIF\_PRE\_FCRATIO} + \varepsilon$$

<i>Independent variable</i>	<i>Status-Matched Firm-Pairs^a</i> <i>(n = 114 firm-pairs)</i>		<i>Quasi-Matched Firm-Pairs^a</i> <i>(n = 149 firm-pairs)</i>	
	<i>Coefficient^b</i>	<i>p-value^c</i>	<i>Coefficient^b</i>	<i>p-value^c</i>
Intercept	0.011	0.059	0.010	0.085
CAPTIVE	0.000	0.998	0.010	0.346
DIF_PRE_SALES	0.008	0.341	0.005	0.531
DIF_PRE_FCRATIO	-0.111	0.523	0.066	0.676
Regression R ²	0.010		0.010	
Intercept (without CAPTIVE) ^d	0.011	0.039	0.014	0.006

^a CAPTIVE is coded 1 if a firm is captive, 0 if the firm is free. The status-matched sample includes only firm-pairs matched on customer concentration (as in BLV). The quasi-matched sample includes the 114 firm-pairs that match on customer concentration plus the 35 firm-pairs in which a captive adopter is matched with a free non-adopter. In either case, CAPTIVE is coded by reference to JIT adopters' statuses.

^b Coefficients displayed for DIF_PRE_SALES are the actual coefficients $\times 10,000$. The model includes DIF_PRE_SALES and DIF_PRE_FCRATIO to control for the significant differences in firm size and cost structure between the JIT and control samples (see Table 2, Panel C).

^c Under BLV's customer concentration hypothesis, captive status (i.e., CAPTIVE = 1) hinders a JIT adopter's post-adoption ROA performance. Thus, negative β_1 supports BLV's hypothesis. Controlling for between-sample differences in size and cost structure, β_0 is the relative ROA change for free adopters; $(\beta_0 + \beta_1)$ is the relative ROA change for captive adopters. Consistent with our directional hypothesis, intercept p-values are from one-tailed tests; other p-values are from two-tailed tests.

^d These intercepts and p-values are from regressions that exclude the CAPTIVE variable.

Median pre-adoption sales for our adopters are \$397.5 million, but only \$108.5 million for BLV's adopters.¹⁷ Second, Table 2, Panel C indicates that our adopters have higher fixed-cost ratios than our non-adopters ($p = 0.002$). In contrast, our analysis of the BLV samples indicates that BLV *non-adopters* have the higher fixed-cost ratios ($p = 0.047$). Thus, if JIT's benefits are greater for large firms and/or firms with large fixed-cost ratios, then our results may differ from BLV's because our adopter sample contains more large firms and firms with high fixed-cost ratios than does BLV's sample.¹⁸

To examine the effect of adopter size on our H1 (ROA) results, we first partitioned our sample and the BLV sample into large-adopter and small-adopter samples using the median pre-adoption sales in BLV as the partition. We find no relative ROA improvement in either of the two small-firm samples ($N = 23$ and 39 in BLV's and our sample, respectively). The median paired difference in change in ROA is -0.010 (-0.011) in BLV's (our) small-adopter sample, and is not significant ($p = 0.477$ and 0.462 , respectively). However, in the

¹⁷ Most of this difference is attributable to the span-of-operations screen. When we employ the screen, our sample includes 89 adopters with median pre-adoption sales of \$165.33 million.

¹⁸ To examine the effects of fixed-cost ratio differences, we conducted analyses similar to those described in this section for adopter size. We concluded from the analyses that fixed-cost ratio differences are not a factor in the two studies' different sample-wide results.

large-firm samples ($N = 23$ and 162 in BLV's and our sample, respectively), the median paired differences are 0.021 in the BLV sample and 0.014 in our sample ($p = 0.031$ and 0.005 , respectively).

Next, we considered whether adopter-size is an important intervening variable in the sample-wide ROA results reported in Table 4. We partitioned our adopter sample on its own (rather than BLV's) median pre-adoption sales value (\$397.5 million) and re-ran our test on the resulting small- and large-adopter samples. The results indicated that both small and large JIT adopters enjoy ROA performance superior to that of non-adopters ($p = 0.052$ and 0.045 , respectively). Further, the correlation between relative ROA improvement and adopter size for our full sample of 201 adopters is insignificant ($p = 0.322$), and multiple regression models yield no evidence of an interaction between JIT adoption and firm size in explaining our sample firms' ROA changes.

Thus, although an adopter-size effect apparently reconciles our results to BLV's, the absence of a significant size effect across our sample of 201 adopters suggests that the effect is not linear, but is instead limited to firms below a firm-size threshold. As noted previously, BLV's 23 largest adopters enjoy significant relative ROA improvement, on average, whereas their 23 smallest adopters do not. Among BLV's 23 smallest adopters, median pre-adoption sales are \$33 million; among their 23 largest adopters, median pre-adoption sales are \$322 million. Hence, comparing the two groups' relative ROA changes is clearly a comparison of very small firms to other firms. When our sample is partitioned on BLV's median sales value, our 39 very small adopters (median sales of \$53 million) do not enjoy significant relative ROA improvement, while the 162 other adopters (median sales \$581 million) do, and the inference regarding adopter size and relative ROA changes is identical. Partitioning our sample on *its own* median sales value again results in two samples with strikingly different median sales values (\$127 million vs. \$1,313 million). Yet, as reported above, both groups enjoy relative ROA gains, evidently because, under this partitioning, the small-adopter group is less dominated by very small firms. Collectively, these results suggest that the difference in inferences between the two studies arises primarily from the much greater presence of very small firms in BLV's sample.¹⁹

Two explanations for the apparent ineffectiveness of JIT for very small adopters seem plausible. First, Alles, et al. (1995, 1996) note that their results "suggest that JIT is more likely to be useful for purposes other than inventory management in circumstances where there are substantial information asymmetries between workers and management." If such information asymmetry is lacking in very small firms, then very small JIT adopters, which constitute a greater proportion of the BLV sample than our sample, are unlikely to realize significant benefits, apart from improved inventory utilization. Second, the adopter-size result is consistent with BLV's suggestion that firms with limited market power may be less able to retain any financial benefits of adopting JIT.

Tests Examining Adoption Timing Effects

We considered whether our inclusion of 78 firms that adopted JIT after 1989 (BLV's latest adoption year) contributes to the two studies' differing ROA results. We partitioned

¹⁹ Of BLV's 46 adopters, 11 (23.9 percent) report pre-adoption sales less than \$21.5 million. In our sample, only 3.5 percent of adopters report pre-adoption sales less than \$21.5 million. We make no claim that our sample, compared to BLV's, is more representative of the underlying population. We constructed a benchmark sample using all Compustat firms in one-digit SICs 2, 3, and 5 that reported data sufficient to calculate BLV's test variables in each year from 1984–1990 (corresponding, approximately, to BLV's data availability requirements). 1,854 Compustat firms met these conditions. Using 1988 (approximately BLV's weighted-average adoption year) as the benchmark year, these 1,854 firms report median sales of \$166.2 million. The median sales value for BLV's 92 sample firms is at the 43rd percentile of the benchmark sample; the median sales value in our sample is at the 61st percentile of the benchmark sample.

our sample into 123 early (pre-1990) and 78 late (post-1989) adopters and re-ran our ROA test. For early adopters, paired differences in ROA changes were significant ($p = 0.007$), but relative ROA changes were not significant for late adopters ($p = 0.218$). Thus, these test results do not reconcile the two studies' results, but instead are more consistent with a "first mover" advantage, and with a scenario in which firms with the most to gain from JIT are among the earliest to adopt.

Adopter Size and Adoption Timing Effects in H2-H4

Finally, we examined the effects of adopter size and adoption timing on the association between JIT adoption and relative changes in asset turnover and profit margin. With the sample partitioned on BLV's median pre-adoption sales, the results mirror those described for ROA. $DIF\Delta ASSETTURN$, $DIF\Delta MARGIN$, and DIF_EFFECT are all reasonably significant in the large-firm sample ($p \leq 0.067$), but insignificant in the small-firm sample ($p \geq 0.164$). With our sample partitioned on its own median pre-adoption sales, $DIF\Delta MARGIN$ is reasonably significant in both samples ($p \leq 0.097$), but $DIF\Delta ASSETTURN$ is significant in the small-adopter sample ($p = 0.049$) and insignificant in the large-adopter sample ($p = 0.214$). Consistent with these results, we also find that DIF_EFFECT is significant for the large-adopter sample ($p = 0.043$), but is insignificant for the small-adopter sample ($p = 0.217$).

In adoption timing tests using ROA components, we find somewhat divergent results for $DIF\Delta ASSETTURN$ and $DIF\Delta MARGIN$. $DIF\Delta ASSETTURN$ is insignificant for early adopters ($p = 0.170$), but significant for late adopters ($p = 0.034$). Conversely, $DIF\Delta MARGIN$ is significant for early adopters ($p = 0.012$), but insignificant for late adopters ($p = 0.230$), and DIF_EFFECT is also significant for early adopters ($p = 0.037$) and insignificant for late adopters ($p = 0.386$).²⁰

VI. CONCLUSION

In this study, we use a large sample of JIT adopters and matched non-adopters to examine the association between JIT adoption and financial performance. Our study makes several contributions to the literature.

First, we find that JIT adopters outperformed matched firms in ROA improvement over a three-year post-adoption period. JIT adopters' pre- to post-adoption ROA changes exceeded those of matched non-adopters by an average of 1.0 percent—a highly significant difference in our primary test. However, the results of additional analyses suggest that very small firms may realize minimal benefits from JIT, a finding that reconciles our inferences with those of BLV, whose sample included a much greater proportion of such firms. The finding that very small JIT adopters achieve no significant relative ROA improvement is consistent with Alles et al.'s (1995) observation that JIT most benefits firms with significant information asymmetry between workers and managers, and with BLV's prediction that a firm's market power may play a significant role in determining its ability to retain JIT-related financial benefits.

The evidence of a sample-wide association between JIT adoption and ROA improvement is from tests incorporating three-year pre- and post-adoption windows. However, the results of additional analyses suggest that relative ROA improvement is concentrated among the earliest JIT adopters (which may have the most to gain from JIT adoption), and that

²⁰ For all four variables of interest ($DIF\Delta ROA$, $DIF\Delta ASSETTURN$, $DIF\Delta MARGIN$, and DIF_EFFECT), multiple regression models estimated with our sample of 201 firm-pairs yield no evidence of statistically significant effects related to adopter size or adoption timing, or to an interaction thereof.

by the fifth or sixth year following JIT adoption, adopters no longer exhibit superior performance in inventory turnover or ROA. Taken together, these results suggest a “first mover” advantage for early adopters, followed by dissipation of the advantage as JIT becomes more widely adopted.

BLV reported post-adoption relative ROA increases only among those JIT adopters not reporting the existence of a major customer under SFAS No. 14. BLV infer that relatively powerful customers expropriate JIT’s benefits from adopters with concentrated customer bases. In contrast, we find no evidence that the realization and retention of JIT-related benefits depends on a diverse customer base.

BLV contend that powerful customers may capture adopters’ JIT-related cost savings. However, JIT requires substantial coordination with upstream suppliers and downstream customers in ordering, production scheduling, and delivery. If downstream coordination costs are increasing in the number of customers with whom JIT adopters must maintain relationships, then “captive” JIT adopters may actually enjoy lower downstream coordination costs. Thus, an SFAS No. 14 disclosure may proxy for two constructs (a firm’s relative market power and its coordination costs) expected to have opposite effects on post-adoption earnings. Future research may test the customer concentration and coordination cost hypotheses, particularly if more suitable proxies for the two constructs can be identified.

Our sample-wide results suggest that JIT adopters, compared to non-adopters, improve both components of ROA—profit margin and asset turnover. However, we extend prior literature by identifying the *primary* source of JIT adopters’ relative ROA increases. Our sample-wide evidence that improved profit margins are the main source of post-JIT adoption ROA improvement (1) is consistent with Alles et al.’s (1995, 188) contention that JIT makes the production setting “transparent,” assisting line workers and management alike in making cost-saving improvements, and (2) indicates that JIT’s benefits are not limited to reduced inventory holding costs and reduced total investment arising from leaner inventories.²¹

Our study has several limitations. Our self-selection bias analysis relies on an untested model of JIT choice. Thus, we cannot rule out the possibility that our reported results stem at least partially from underlying dissimilarities between JIT adopters and non-adopters. We leave to future research more sophisticated modeling of JIT choice and further study of the potential endogeneity in the JIT adoption decision. The interdependence of profit margin and asset turnover increases the difficulty of assessing the two measures’ relative importance in the association between JIT adoption and ROA performance. Thus, our analyses regarding the source of the positive association between JIT adoption and ROA performance are only an initial attempt to refine our understanding of JIT’s performance effects. Finally, our study covers a period in which American companies adopted a number of initiatives aside from JIT, such as cellular manufacturing, activity-based costing, and total quality management. Concurrent adoptions of JIT and other initiatives require that readers exercise caution when interpreting our study’s results.

²¹ We cannot definitively attribute our H4 findings to either reduced inventory holding costs or reduced production costs. It seems unlikely, however, that reduced holding costs alone would drive our sample-wide finding that profit margin dominates asset turnover in explaining JIT adopters’ relatively improved ROAs.

REFERENCES

- Alles, M., S. M. Datar, and R. A. Lambert. 1995. Moral hazard and management control in just-in-time settings. *Journal of Accounting Research* 33: 177–204.
- Anyane-Ntow, K. 1991. Just-in-time manufacturing systems and inventory reported in financial statements: A cross-national comparison of manufacturing firms. *The International Journal of Accounting* 26: 277–285.

- Atkinson, A. A., R. D. Banker, R. S. Kaplan, and S. M. Young. 2001. *Management Accounting*. 3rd edition. Upper Saddle River, NJ: Prentice Hall.
- Balakrishnan, R., T. J. Linsmeier, and M. Venkatachalam. 1996. Financial benefits from JIT adoption: Effects of customer concentration and cost structure. *The Accounting Review* 71 (April): 183–205.
- Biggart, T. B. 2000. The effects of just-in-time inventory system adoption on firm performance. Working paper, University of North Carolina at Greensboro.
- Callen, J. L., C. Fader, and I. Krinsky. 2000. Just-in-time: A cross-sectional plant analysis. *International Journal of Production Economics* 63: 277–301.
- Fullerton, R. R., and C. S. McWatters. 2001. An empirical investigation of the relationship between JIT and financial performance. Working paper, Utah State University.
- Hogan, C. E. 1997. Costs and benefits of audit quality in the IPO market: A self-selection analysis. *The Accounting Review* (January): 67–86.
- Huson, M., and D. Nanda. 1995. The impact of just-in-time manufacturing on firm performance in the U.S. *Journal of Operations Management* 12: 297–310.
- Lubben, R. T. 1988. *Just-In-Time Manufacturing: An Aggressive Manufacturing Strategy*. New York, NY: McGraw-Hill.
- Maddala, G. S. 1977. Self-selectivity problems in econometric models. In *Applications of Statistics*, edited by P. R. Krishnaiah. New York, NY: Elsevier Science.
- . 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs in Quantitative Economics No. 3. New York, NY: Cambridge University Press.
- . 1991. A perspective on the use of limited-dependent and qualitative variables models in accounting research. *The Accounting Review* 66 (October): 788–807.
- McIlhatten, R. D. 1987. How cost management can support the JIT philosophy. *Management Accounting* 69 (September): 20–27.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817–838.
- Young, S. M., and F. H. Selto. 1991. New manufacturing practices and cost management: A review of the literature and directions for new research. *Journal of Accounting Literature* 10: 265–297.

BOOK REVIEWS

Jane L. Reimers, Editor

Editor's Note: The book reviews in this issue address very current topics in accounting, and they provide ideas and insights useful for both teaching and research. Thanks to both of the reviewers for their comments on these two recently published books.

PRICEWATERHOUSECOOPERS LLP and SAP AG, *The E-Business Workplace: Discovering the Power of Enterprise Portals* (New York, NY: John Wiley & Sons, Inc., 2001, pp. 244, \$49.95 Hb).

For information systems educators of the 1970s and early 1980s, who can forget the advantages that were touted to accrue to the entities that were able to fully integrate a "management information system"? The initial concept of using one database to meet all the needs of management was viewed as unprecedented. *The E-Business Workplace* surpasses that concept to suggest not only that the entity must look far beyond its walls for the appropriate information, but also that an entity must develop a delivery system that focuses on each employee's role in the organization.

Promoted is the concept that a "workplace," or an enterprise portal tied to the Web, should be developed to provide each member of an entity with access to information, applications, and services necessary to fulfill his or her role. Rather than being tied to a physical work environment, the workplace would be accessible via more traditional PC browser technology, as well as personal digital assistants (PDAs), cell phones, and other technologies not yet envisioned. The workplace concept is not restricted to insiders of the organization. The concept is envisioned that a workplace environment would allow interaction among all members of the organization's stakeholders.

The E-Business Workplace is a collaborative effort by a team of 27 authors of PricewaterhouseCoopers LLP and SAP AG. To the credit of the authors, the delivery is smooth and the text is not a sales pitch for either PWC or SAP. However, the CD-ROM included with the text is punctuated by references to SAP's Enterprise Resource Planning (ERP) software and uses SAP examples to demonstrate concepts.

The audience for the book would be managers of organizations attempting to exploit the advantages of e-commerce in the work environment. It should cause managers and any other decision maker to rethink the acquisition and use of information. The book would not be appropriate for an undergraduate course, but may have some benefit as a supplemental reading for a graduate course in information systems.

The first part of the book is devoted to an appeal for managers to think beyond "integration" of information to "collaboration" of information with the stakeholders of the entity. Advantages of such an approach are detailed. The balance of the book addresses the issues of implementation of a workplace environment. The steps for development are outlined, and the various risks are considered. Examples are interspersed into the text (as well as into the CD-ROM).

One of the strong points of the text is the focus on change. It notes that knowledge must be managed like other assets and that knowledge needs are dynamic. Managers must ensure that individuals "buy in to change" and that they recognize the benefits. Obviously, training is an important component of the workplace concept. Harmonious with the workplace philosophy is that it can be tailored to deploy web-based training and distance learning as part of an overall training program.

In summary, the book invites the manager to explore the benefits of a collaborative view of information management. It is thought provoking and insightful.

WILLIAM A. HILLISON

Florida State University

YOUNG, S. DAVID, and STEPHEN F. O'BYRNE, *EVA® and Value-Based Management: A Practical Guide to Implementation* (McGraw Hill, 2001, pp. 493).

As the title suggests, this book provides an overview of economic value added (EVA) as well as the broader subject of value-based management. An important feature of the book, not expressed by its title, is that it brings together and distinguishes among various residual income and cash flow concepts, including EVA, refined EVA (REVA), return on net assets (RONA), economic profit, and cash flow return on investment (CFROI). The book gives extensive implementation guidance in the form of designing and implementation of management compensation systems linked to these metrics, and it provides detailed case studies showing how to calculate these metrics for large and complex firms.

There are at least three aspects of this book that I find noteworthy. The first is the thoughtful and integrative discussion of management compensation issues from a value-based management perspective. Chapter 4, for example, does a very nice job of describing the features of the classic principal agent conflict in diffusely held organizations, and the authors highlight why aligning manager and shareholder interests is so important in these settings. EVA is highlighted as one mechanism to achieve such congruence. In addition to providing clear intuition and real-life examples, this chapter also links academic research on management buy-outs, equity carve-outs, and tracking stock to the question of how best to align shareholder and manager interests. The authors revisit issues of management compensation in Chapter 8, where they consider EVA bonus formulae in detail, detailing why changes have been made, over time, to the basic EVA plan.

A second and related (to management compensation) noteworthy aspect of this book is the discussion of "wealth leverage" and the use of stock options as alternatives, or additions, to EVA-based bonus mechanisms. Wealth leverage is defined as the ratio of the percentage change in the manager's wealth to the percentage change in shareholder wealth. A manager whose entire wealth is represented by common stock of his employer, therefore, has wealth leverage of 100 percent, because any change in the stock price is fully experienced by the manager. The authors also devote attention to the question of how to think about stock options as alternatives or additions to EVA-type compensation plans.

The final aspect of this book that I find interesting and noteworthy is Chapters 5 to 10, which are devoted to explaining the implementation of EVA measurement and performance systems. In this regard, Chapter 5 provides a particularly lucid and complete description of approaches to estimating the cost of capital (with special emphasis on using the capital asset pricing model and the arbitrage pricing model), at the same time that it summarizes and evaluates the evidence on concerns with these approaches. Chapter 6 is equally impressive in its descriptions of the key accounting adjustments made to calculate EVA, including adjustments for research and development, deferred taxes, operating leases, goodwill, and restructuring charges. Not only do the authors review the technical details of the accounting adjustments, but they also discuss the economic rationale for making them.

This book is relevant to a broad constituency, including instructors of courses in which residual income valuation or performance evaluation systems are taught, and practitioners interested in a broad overview of EVA, as well as those who want a more detailed discussion of EVA and value-based management systems. The authors cover all of the major issues concerning residual income and cash flow metrics; they clearly describe disputes and contentious points concerning their comparative use, and they provide more substantive and in-depth discussions of implementation issues (such as accounting adjustments) than I have read elsewhere. If I were to sum up my thoughts on this book in one sentence, then it would be this: If you have time to read but one book on EVA, read this one.

JENNIFER FRANCIS

Duke University

Editorial Data

The following table contains information about turnaround time for manuscripts (including revisions) on which editorial decisions were made in the 12-month period ended August 31, 2001. Turnaround time is the number of days between the date that the manuscript was received *and the date of the editor's letter to the author(s)*.

	<u>Number of Manuscripts</u>	<u>Cumulative Number</u>	<u>Cumulative Percent</u>
0 ≤ Days ≤ 30	78	78	25.16
31 ≤ Days ≤ 60	115	193	62.26
61 ≤ Days ≤ 90	67	260	83.55
91 ≤ Days ≤ 120	34	294	98.06
121 ≤ Days	16	310	100.00

The mean review time was 55 days; the median review time was 51 days.

OCCASIONAL REVIEWERS AND ASSOCIATE EDITORS

The following faculty have graciously donated their time and expertise to assist in evaluating manuscripts for *The Accounting Review* between June 2000 and May 2001. Their support is greatly appreciated.—*The Editor*

David Aboody, University of California, Los Angeles

Anwer S. Ahmed, Syracuse University

Eli Amir, Tel Aviv University

Regina M. Anctil, University of Minnesota

Mark C. Anderson, The University of Texas at Dallas

Anil Arya, The Ohio State University

Hollis Ashbaugh, University of Wisconsin—Madison

Andrew D. Bailey, Jr., University of Illinois at Urbana—Champaign

Tim Baldenius, Columbia University

Chip Bamberger, Lexecon Inc.

Sati P. Bandyopadhyay, University of Waterloo

Ronnie Barnes, London Business School

Sudipta Basu, Emory University

Mark S. Beasley, North Carolina State University

Jean Bedard, Northeastern University

Joy Begley, University of British Columbia

Shlomo Benartzi, University of California, Los Angeles

Messod D. Beneish, Indiana University

Nilabhra Bhattacharya, University of Utah

A. Faye Borthick, Georgia State University

Christine A. Botosan, University of Utah

Stephen Brown, Emory University

Donal Byard, University of Cincinnati

Jeffrey L. Callen, University of Toronto

Dennis H. Caplan, Iowa State University

Joseph V. Carcello, University of Tennessee

Derek Chan, Hong Kong University of Science and Technology

Byeonghee Choi, McGill University

Bryan K. Church, Georgia Institute of Technology

Julie H. Collins, The University of North Carolina at Chapel Hill

John Core, University of Pennsylvania

Donald P. Cram, California State University, Fullerton

William M. Cready, Louisiana State University

Antonio Davila, Stanford University

Elizabeth B. Davis, Baylor University

Paquita Y. Davis-Friday, University of Notre Dame

F. Todd DeZoort, The University of Alabama

Bala G. Dharan, Rice University

Mark W. Dirmsmith, Pennsylvania State University

Ronald A. Dye, Northwestern University

John W. Eichenhofer, University of Wisconsin

Leslie G. Eldenburg, The University of Arizona

Kirsten M. Ely, Georgia Institute of Technology

Ellen Engel, University of Chicago

Patricia M. Fairfield, Georgetown University

Gerald A. Feltham, University of British Columbia

Martin G. Fennema, Florida State University

George Foster, Stanford University

Jennifer Francis, Duke University

Richard M. Frankel, Massachusetts Institute of Technology

Veronique G. Frucot, Christopher Newport University

Ananda R. Ganguly, University of Illinois

Jonathan C. Glover, Carnegie Mellon University

Brian P. Green, University of Michigan—Dearborn

J. Douglas Hanna, University of Chicago

Stephen C. Hansen, University of California, Los Angeles

David Haushalter, University of Oregon

Rachel M. Hayes, University of Chicago

Carla K. Hayn, University of California, Los Angeles

Thomas Hemmer, University of Chicago

William G. Heninger, Brigham Young University

Dana R. Hermanson, Kennesaw State University

Stephen A. Hillegeist, Northwestern University

Joanna L. Ho, University of California, Irvine

Cameron Hooper, University of Michigan

Richard W. Houston, The University of Alabama

S. Paul Hribar, Cornell University

- Mark Huson, University of Alberta
 Lee-Seok Hwang, Baruch College–CUNY
 Prem C. Jain, Georgetown University
 Marilyn F. Johnson, Michigan State University
 Karla M. Johnstone, University of Wisconsin–Madison
 Woon-Oh Jung, Seoul National University
 Sanjay Kallapur, Purdue University
 Sok-Hyon Kang, The George Washington University
 Jonathan Karpoff, University of Washington
 Ron Kasznik, Stanford University
 Beth B. Kern, Indiana University South Bend
 Kimberly G. Key, Auburn University
 Ronald R. King, Washington University
 Michael Kirschenheiter, Columbia University
 April Klein, New York University
 W. Robert Knechel, University of Florida
 Adam S. Koch, Carnegie Mellon University
 Jagan Krishnan, Temple University
 Ranjani Krishnan, Michigan State University
 Krishna R. Kumar, The George Washington University
 Richard A. Lambert, University of Pennsylvania
 Wayne R. Landsman, The University of North Carolina at Chapel Hill
 Reuven Leavy, University of California, Berkeley
 Andrew J. Leone, University of Rochester
 Carolyn B. Levine, Carnegie Mellon University
 Frederick W. Lindahl, The George Washington University
 Thomas J. Linsmeier, Michigan State University
 Marc Lipson, University of Georgia
 Kin Lo, The University of British Columbia
 Steven Lustgarten, Baruch College–CUNY
 Luann Lynch, University of Virginia
 Robert P. Magee, Northwestern University
 Edward L. Maydew, The University of North Carolina at Chapel Hill
 James C. McKeown, Pennsylvania State University
 William F. Messier, Jr., Georgia State University
 Michael Mikhail, Duke University
 H. Fred Mittelstaedt, University of Notre Dame
 Cindy L. Moeckel, Arizona State University
 Steve Monahan, The University of Chicago
 Donald V. Moser, University of Pittsburgh
 Kevin E. Murphy, University of Southern California
 Jane F. Mutchler, Georgia State University
 Venky Nagar, University of Michigan
 Dhananjay Nanda, University of Michigan
 Ramachandran Natarajan, The University of Texas at Dallas
 Siva Nathan, Georgia State University
 Karen K. Nelson, Stanford University
 Kaye J. Newberry, The University of Arizona
 Terrance T. Odean, University of California, Davis
 Edmund Outslay, Michigan State University
 Neil Pearson, University of Illinois at Urbana–Champaign
 Stephen H. Penman, Columbia University
 Mark C. Penno, Purdue University
 Ray J. Pfeiffer, University of Massachusetts
 Morton Pincus, The University of Iowa
 George A. Plesko, Massachusetts Institute of Technology
 C. Elizabeth Plummer, Southern Methodist University
 Srinivasan Rangan, University of Colorado
 Graeme Rankine, The American Graduate School of International Management
 Philip M. J. Reckers, Arizona State University
 Shelley Rhoades-Catanach, Villanova University
 David N. Ricchiute, University of Notre Dame
 Steven E. Salterio, University of Waterloo
 Bharat Sarath, Baruch College–CUNY
 Michael S. Schadeewald, University of Wisconsin–Milwaukee
 Jeffrey W. Schatzberg, The University of Arizona
 Catherine M. Schrand, University of Pennsylvania
 Rachel Schwartz, Washington University
 Steven T. Schwartz, SUNY at Binghamton
 Thomas W. Scott, University of Alberta
 John T. Sennetti, Nova Southeastern University
 Galen R. Sevcik, Georgia State University
 Brian P. Shapiro, University of Minnesota
 Kenneth W. Shaw, University of Maryland
 Aamer Sheikh, College of William & Mary
 Daniel T. Simon, University of Notre Dame
 Dan A. Simunic, University of British Columbia
 Kumar N. Sivakumar, Boston University
 Konduro Sivaramakrishnan, Texas A&M University

Douglas J. Skinner, University of Michigan
Michael Smith, Duke University
Naomi S. Soderstrom, University of Colorado
Billy S. Soo, Boston College
Theodore Sougiannis, University of Illinois
Andrew E. Spero, University of Minnesota
Brian C. Spilker, Brigham Young University
Eric E. Spires, The Ohio State University
Sri S. Sridhar, Northwestern University
Dhinu Srinivasan, University of Pittsburgh
Earl K. Stice, Brigham Young University
Clyde P. Stickney, Dartmouth College
Toby Stock, Ohio University
Phillip Stocken, University of Pennsylvania
Gregory M. Trompeter, Boston College
Ken T. Trotman, University of New South
Wales
Peter Tufano, Harvard University
Mohan Venkatachalam, Stanford University
Elizabeth Venuti, Hofstra University
Sandra C. Vera-Muñoz, University of Notre
Dame

Gnanakumar Visvanathan, The George
Washington University
James M. Wahlen, Indiana University
Beverly R. Walther, Northwestern University
Michael Welker, Queen's University
Christine I. Wiedman, University of Western
Ontario
Michael Willenborg, University of Connecticut
David D. Williams, The Ohio State University
Richard Willis, Duke University
Earl R. Wilson, University of Missouri
G. Peter Wilson, Boston College
Martin G. H. Wu, University of Illinois
Peter D. Wysocki, Massachusetts Institute of
Technology
Sung-Soo Yoon, University of California, Los
Angeles
Richard A. Young, The Ohio State University
Susan M. Young, Emory University
Xiao-Jun Zhang, University of California,
Berkeley

EDITORIAL POLICY AND STYLE INFORMATION

EDITORIAL POLICY

According to the policies set by the Publications Committee (which were endorsed by the Executive Committee and were published in the *Accounting Education News*, June 1987), *The Accounting Review* “should be viewed as the premier journal for publishing articles reporting the results of accounting research and explaining and illustrating related research methodology. The scope of acceptable articles should embrace any research methodology and any accounting-related subject, as long as the articles meet the standards established for publication in the journal....No special sections should be necessary. The primary, but not exclusive, audience should be—as it is now—academicians, graduate students, and others interested in accounting research.”

The primary criterion for publication in *The Accounting Review* is the significance of the contribution an article makes to the literature.

The efficiency and effectiveness of the editorial review process is critically dependent upon the actions of both authors submitting papers and the reviewers. Authors accept the responsibility of preparing research papers at a level suitable for evaluation by independent reviewers. Such preparation, therefore, should include subjecting the manuscript to critique by colleagues and others and revising it accordingly prior to submission. The review process is not to be used as a means of obtaining feedback at early stages of developing the research.

Reviewers and associate editors are responsible for providing critically constructive and prompt evaluations of submitted research papers based on the significance of their contribution and on the rigor of analysis and presentation. Associate editors also make editorial recommendations to the editor.

MANUSCRIPT PREPARATION AND STYLE

The Accounting Review's manuscript preparation guidelines follow (with a slight modification) the B-format of *The Chicago Manual of Style* (14th ed.; University of Chicago Press). Another helpful guide to usage and style is *The Elements of Style*, by William Strunk, Jr., and E. B. White (Macmillan). Spelling follows *Webster's International Dictionary*.

FORMAT

1. All manuscripts should be typed in 12-point font on one side of 8½ × 11" good quality paper and be double spaced, except for indented quotations.
2. Manuscripts should be as concise as the subject and research method permit, generally not to exceed 7,000 words.
3. Margins of at least one inch from top, bottom, and sides should facilitate editing and duplication.
4. To promote anonymous review, authors should not identify themselves directly or indirectly in their papers or in experimental test instruments included with the submission. Single authors should not use the editorial “we.”
5. A cover page should show the title of the paper, the author's name, title and affiliation, email address, any acknowledgments, and a footnote indicating whether the author would be willing to share the data (see last paragraph in this statement).

Pagination: All pages, including tables, appendices and references, should be serially numbered. Major sections should be numbered in Roman numerals. Subsections should not be numbered.

Numbers: Spell out numbers from one to ten, except when used in tables and lists, and when used with mathematical, statistical, scientific, or technical units and quantities, such as distances, weights and measures. For example: *three days*; *3 kilometers*; *30 years*. All other numbers are expressed numerically.

Percentages and Decimal Fractions: In nontechnical copy use the word percent in the text.

Hyphens: Use a hyphen to join unit modifiers or to clarify usage. For example: *a well-presented analysis*; *re-form*. See *Webster's* for correct usage.

Keywords: The abstract is to be followed by four keywords that will assist in indexing the paper.

ABSTRACT/INTRODUCTION

An Abstract of about 100 words should be presented on a separate page immediately preceding the text. The Abstract should concisely inform the reader of the manuscript's topic, its methods, and its findings.

Keywords and the Data Availability statements should follow the Abstract. The text of the paper should start with a section labeled "I. Introduction," which provides more details about the paper's purpose, motivation, methodology, and findings. Both the Abstract and the introduction should be relatively nontechnical, yet clear enough for an informed reader to understand the manuscript's contribution. The manuscript's title, but neither the author's name nor other identification designations, should appear on the Abstract page.

TABLES AND FIGURES

The author should note the following general requirements:

1. Each table and figure (graphic) should appear on a separate page and should be placed at the end of the text. Each should bear an Arabic number and a complete title indicating the exact contents of the table or figure. Tables and figures should define each variable. The titles and definitions should be sufficiently detailed to enable the reader to interpret the tables and figures without reference to the text.
2. A reference to each graphic should be made in the text.
3. The author should indicate by marginal notation where each graphic should be inserted in the text.
4. Graphics should be reasonably interpreted without reference to the text.
5. Source lines and notes should be included as necessary.

Equations: Equations should be numbered in parentheses flush with the right-hand margin.

DOCUMENTATION

Citations: Work cited should use the "author-date system" keyed to a list of works in the reference list (see below). Authors should make an effort to include the relevant page numbers in the cited works.

1. In the text, works are cited as follows: authors' last name and date, without comma, in parentheses: for example, (Jones 1987); with two authors: (Jones and Freeman 1973); with more than two: (Jones et al. 1985); with more than one source cited together (Jones 1987; Freeman 1986); with two or more works by one author: (Jones 1985, 1987).
2. Unless confusion would result, do not use "p." or "pp." before page numbers: for example, (Jones 1987, 115).
3. When the reference list contains more than one work of an author published in the same year, the suffix a, b, etc. follows the date in the text citation: for example, (Jones 1987a) or (Jones 1987a; Freeman 1985b).
4. If an author's name is mentioned in the text, it need not be repeated in the citation; for example, "Jones (1987, 115) says...."
5. Citations to institutional works should use acronyms or short titles where practicable; for example, (AAA ASOBAT 1966); (AICPA Cohen Commission Report 1977). Where brief, the full title of an institutional work might be shown in a citation: for example, (ICAEW *The Corporate Report* 1975).
6. If the manuscript refers to statutes, legal treatises or court cases, citations acceptable in law reviews should be used.

Reference List: Every manuscript must include a list of references containing only those works cited. Each entry should contain all data necessary for unambiguous identification. With the author-date system, use the following format recommended by *The Chicago Manual*:

1. Arrange citations in alphabetical order according to surname of the first author or the name of the institution responsible for the citation.
2. Use author's initials instead of proper names.
3. Date of publication should be placed immediately after author's name.
4. Titles of journals should not be abbreviated.
5. Multiple works by the same author(s) in the same year are distinguished by letters after the date.
6. Inclusive page numbers are treated as recommended in *Chicago Manual* section 8.67.

Sample entries are as follows:

- American Accounting Association, Committee on Concepts and Standards for External Financial Reports. 1977. *Statement on Accounting Theory and Theory Acceptance*. Sarasota, FL: AAA.
- Demski, J. S., and D. E. M. Sappington. 1989. Hierarchical structure and responsibility accounting. *Journal of Accounting Research* 27 (Spring): 40-58.
- Dye, R., B. Balachandran, and R. Magee. 1989. Contingent fees for audit firms. Working paper, Northwestern University, Evanston, IL.

- Fabozzi, F., and I. Pollack, eds. 1987. *The Handbook of Fixed Income Securities*. 2nd edition. Homewood, IL: Dow Jones-Irwin.
- Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, U.K.: Cambridge University Press.
- Porcano, T. M. 1984a. Distributive justice and tax policy. *The Accounting Review* 59 (October): 619–636.
- . 1984b. The perceived effects of tax policy on corporate investment intentions. *The Journal of the American Taxation Association* 6 (Fall): 7–19.
- Shaw, W. H. 1985. Empirical evidence on the market impact of the safe harbor leasing law. Ph.D. dissertation, The University of Texas at Austin.
- Sherman, T. M., ed. 1984. *Conceptual Framework for Financial Accounting*. Cambridge, MA: Harvard Business School.

Footnotes: Footnotes are not used for documentation. Textual footnotes should be used only for extensions and useful excursions of information that if included in the body of the text might disrupt its continuity. Footnotes should be consecutively numbered throughout the manuscript with superscript Arabic numerals. Footnote text should be doubled-spaced and placed at the end of the article.

SUBMISSION OF MANUSCRIPTS

Authors should note the following guidelines for submitting manuscripts:

1. Manuscripts currently under consideration by another journal or publisher should not be submitted. The author must state that the work is not submitted or published elsewhere.
2. In the case of manuscripts reporting on field surveys or experiments, four copies of the instrument (questionnaire, case, interview plan or the like) should be submitted. Information that might identify the author(s) must be deleted from the instrument.
3. Four copies should be submitted together with a check in U.S. funds for \$75.00 for members or \$100.00 for nonmembers of the AAA made payable to the American Accounting Association. Effective January 1990, the submission fee is nonrefundable.
4. The author should retain a copy of the paper.
5. Revisions must be submitted within 12 months from request, otherwise they will be considered new submissions.

COMMENTS

Comments on articles previously published in *The Accounting Review* will be reviewed (anonymously) by two reviewers in sequence. The first reviewer will be the author of the original article being subjected to critique. If substance permits, a suitably revised comment will be sent to a second reviewer to determine its publishability in *The Accounting Review*. If a comment is accepted for publication, the original author will be invited to reply. All other editorial requirements, as enumerated above, also apply to proposed comments.

POLICY ON REPRODUCTION

An objective of *The Accounting Review* is to promote the wide dissemination of the results of systematic scholarly inquiries into the broad field of accounting.

Permission is hereby granted to reproduce any of the contents of the *Review* for use in courses of instruction, as long as the source and American Accounting Association copyright are indicated in any such reproductions.

Written application must be made to the American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399, for permission to reproduce any of the contents of the *Review* for use in other than courses of instruction—e.g., inclusion in books of readings or in any other publications intended for general distribution. In consideration for the grant of permission by the *Review* in such instances, the applicant must notify the author(s) in writing of the intended use to be made of each reproduction. Normally, the *Review* will not assess a charge for the waiver of copyright.

Except where otherwise noted in articles, the copyright interest has been transferred to the American Accounting Association. Where the author(s) has (have) not transferred the copyright to the Association, applicants must seek permission to reproduce (for all purposes) directly from the author(s).

POLICY ON DATA AVAILABILITY

The following policy has been adopted by the Executive Committee in its April 1989 meeting.

"An objective of (*The Accounting Review, Accounting Horizons, Issues in Accounting Education*) is to provide the widest possible dissemination of knowledge based on systematic scholarly inquiries into accounting as a field of professional research, and educational activity. As part of this process, authors are encouraged to make their data available for use by others in extending or replicating results reported in their articles. Authors of articles which report data dependent results should footnote the status of data availability and, when pertinent, this should be accompanied by information on how the data may be obtained."

Accounting Conservatism, the Quality of Earnings, and Stock Returns

Stephen H. Penman

Columbia University

Xiao-Jun Zhang

University of California, Berkeley

ABSTRACT: When a firm practices conservative accounting, changes in the amount of its investments can affect the quality of its earnings. Growth in investment reduces reported earnings and creates reserves. Reducing investment releases those reserves, increasing earnings. If the change in investment is temporary, then current earnings is temporarily depressed or inflated, and thus is not a good indicator of future earnings. This study develops diagnostic measures of this joint effect of investment and conservative accounting. We find that these measures forecast differences in future return on net operating assets relative to current return on net operating assets. Moreover, these measures also forecast stock returns—indicating that investors do not appreciate how conservatism and changes in investment combine to raise questions about the quality of reported earnings.

Keywords: *earnings quality; conservative accounting; forecasting accounting rates of return; predicting stock returns.*

Data Availability: *Data are available from sources identified in the paper.*

I. INTRODUCTION

This paper contributes to the stream of research on how accounting methods affect the quality of earnings. While the academic and professional literature has not reached a consensus on the definition of *quality of earnings*, we define the term to mean that reported earnings, before extraordinary items that are readily identified on the income statement, is of good quality if it is a good indicator of future earnings. Thus we

The comments of participants in workshops at Barclays Global Investors, Baruch College of CUNY, Columbia University, Cornell University, Georgetown University, the University of Michigan Quality of Earnings Conference, University of Minnesota, and Stanford University are appreciated, particularly those of Ilia Dichev, Tom Dyckman, John Elliott, Patricia Fairfield, Teri Yohn, and Paul Zarowin. We also appreciate comments of Mary Barth (associate editor) and anonymous reviewers.

*Submitted December 1999
Accepted September 2001*

consider high-quality earnings to be “sustainable earnings,” as often referred to in financial analysis (in Mulford and Comiskey [1996], for example). Correspondingly, when an accounting treatment produces unsustainable earnings, we deem those unsustainable earnings to be of poor quality.

Most of the research on the quality of earnings focuses on the effects of changes in accounting estimates (see Healy and Wahlen [1999] for a recent review). Estimates are applied in the accruals, so in order to identify poor-quality earnings, papers have investigated total accruals (as in Healy 1985), discretionary accruals (as in Jones 1991), and extreme accruals (as in Sloan 1996). This research characterizes managers as creating unsustainable earnings, for example, by temporarily reducing estimates of bad debts or depreciation (as in Teoh et al. 1998) or lowering loan loss reserves (as in Beaver and Engel 1996)—or lowering current earnings (and increasing future earnings) by overestimating restructuring charges (as in Moehrle 2002). If the effect is temporary, and so reverses later through a revision in the estimate, then reported earnings are of poor quality because current earnings are not a good indicator of future earnings. Under this view, applying accounting principles consistently from period to period leads to more sustainable earnings, and so reduces quality concerns.

This paper shows empirically that quality concerns can also arise if firms apply conservative accounting consistently without any change in accounting methods or estimates.

By *conservative accounting* we mean choosing accounting methods and estimates that keep the book values of net assets relatively low. Therefore, LIFO accounting for inventories is conservative relative to FIFO (if inventory costs are increasing); expensing research and development (R&D) expenditures rather than capitalizing and amortizing them is conservative; depreciation methods that consistently use short estimated asset lives (so as to record depreciation in excess of economic depreciation) are conservative; and policies that consistently overestimate allowances for doubtful accounts, sales returns, or warranty liabilities are conservative.

Conservative accounting affects not only the quality of the numbers reported on the balance sheet, but also the quality of earnings reported on the income statement.¹ When the firm increases investment, conservative accounting leads to reported earnings that are indeed lower than they would have been had management made more liberal accounting choices. These lower earnings, however, create unrecorded reserves that provide managers with flexibility to report more income in the future. Management can increase these reserves, and so reduce earnings, by increasing investment. Management can also release the reserves and create additional earnings, by subsequently reducing investment or reducing the rate of growth in investment. While the effect of the interaction of conservatism and investment is a mechanical one, it is not necessarily temporary; the temporary extent of the effect has to be ascertained. If the change in investment is temporary, then the induced change in earnings is also temporary and not indicative of subsequent earnings. Hence, the quality of earnings depends on the interaction between real activity and accounting policy, not merely on changes in accounting methods and estimates. These arguments suggest that managers can use the joint effect of real activity and accounting policy to manage earnings. However, the effect arises, intended or not. The effect is perverse—reducing investment

¹ Indeed, earnings quality issues are interesting only when balance sheets are imperfect indicators of value. When balance sheets are marked-to-market so that book value is equal to value, earnings are of low quality as an indicator of future earnings; earnings are, in this case, simply changes in value, and changes in value are uninformative about future earnings. Imperfect balance sheets shift the focus to the income statement as an indicator of value. The quality of earnings therefore becomes more important as an indicator of the future earnings, and thus of value. That quality does, however, depend on how the accountant measures the balance sheet, as explained below.

reduces future earnings but, with conservative accounting, increases current earnings, making current earnings a poor indicator of future earnings.

We pose two research questions. First, does conservative accounting coupled with changes in investment lead to temporary changes in earnings? Second, does the stock market price these temporary earnings as if they were indeed unsustainable? The answer to the first question appears to be yes; the answer to the second is no. In carrying out the analysis, we develop metrics designed to assess whether the effect on earnings is temporary, and we suggest that these metrics can serve as tools for financial analysis and research into the quality of earnings.

Additional investment typically yields additional future earnings; therefore, to examine the sustainability of earnings, we focus on accounting rates of return—earnings relative to net assets that deflates earnings for new net investment. Focusing on accounting rates of return also serves our purpose of investigating the stock market's pricing of earnings of varying quality. Accrual-accounting residual income valuation models dictate that forecasted earnings are evaluated relative to the book values that generate them. Residual income models therefore describe intrinsic equity values as determined by anticipated accounting rates of return. Pricing errors occur if the market, relying on an unsustainable current rate of return, forecasts future rates of return incorrectly.

Prior research has referred to this view of market inefficiency as “fixation” on reported earnings. In those terms, our evidence suggests that the market fixates on reported earnings, unaware that conservative accounting may lead to reported earnings of doubtful quality. Our paper views the analysis of earnings quality—and the identification of market fixation—as a matter of financial statement analysis that uncovers information in the financial statements (beyond current earnings) about future earnings and equity values. Our study therefore contributes to the emerging empirical research on financial statement analysis and fundamental analysis, such as Ou and Penman (1989), Sloan (1996), Penman and Sougianis (1998), Lee et al. (1999), and Nissim and Penman (2001).

The paper is organized as follows. As a background to the analysis, the next section outlines the effects of conservative accounting and investment growth on accounting rates of return. Section III develops metrics for scoring firms on the degree of their conservative accounting and the quality of their earnings, and Section IV summarizes the data. Sections V and VI address the two research questions; Section V investigates whether we can use the quality scores that capture the effect of conservative accounting and investment to identify the sustainability of accounting rates of return, and Section VI documents the returns to taking positions in stocks based on the earnings quality scores. Section VII summarizes the study's conclusions.

II. CONSERVATISM, INVESTMENT, ACCOUNTING RATES OF RETURN, AND THE QUALITY OF EARNINGS

Definition of Conservatism

Researchers have introduced a variety of definitions of conservative accounting. Some, such as Basu (1997), define conservatism as the practice of reducing earnings (and writing down net assets) in response to “bad news,” but not increasing earnings (and writing up net assets) in response to “good news.” In the accounting-based valuation literature, researchers often refer to Feltham and Ohlson's (1995) characterization of conservative or “biased” accounting as an expectation that reported net assets will be less than market value in the long run. That definition classifies the accounting for anticipated positive net-present-value investments at historical cost as conservative accounting, because the analyst expects those investments to be carried at less than their value. Zhang (2000) models

conservative accounting and valuation from this perspective. Beaver and Ryan (2000) similarly characterize conservatism (or bias) as a persistent difference between market value and book value that is distinct from temporary differences due to economic gains and losses that are recognized in book value gradually over time. Gjesdal (1999) distinguishes "economic profitability" from accounting profitability. He characterizes accounting as conservative if it assigns investments a carrying value that yields an expected accounting rate of return greater than the internal rate of return on their cost. Thus, for example, conservative accounting carries an asset whose value is equal to its historical cost (a zero net present value investment) at less than historical cost, as is the case with the immediate expensing of R&D and advertising costs.

Our notion of conservatism follows the latter definition; that is, conservative accounting is a biased application of historical cost accounting.² For our purposes, however, we need not establish unbiased historical-cost carrying values of investments or unbiased allocations of investment cost to match against revenues. Rather, we examine conservatism in a relative sense. We consider one accounting practice (e.g., LIFO accounting) more conservative than a second accounting practice (FIFO accounting) if the first yields lower accumulated earnings (because of the relatively accelerated expensing of costs) and consequently lower carrying values than the second.

Effects of Conservatism and Investment on Rates of Return

The effect of conservative accounting on accounting rates of return is clear. If net assets do not change over a period, then conservative accounting has no effect on earnings in the numerator of a rate of return. (It is always the case, for example, that earnings are the same under accelerated depreciation and straight-line depreciation if property, plant, and equipment is unchanged over the period.) Because conservative accounting carries assets in the denominator at lower amounts, the rate of return is higher, however. Greenball (1969) shows that conservative accounting (bias) interacts with growth; conservative accounting reduces earnings when there is growth in net assets so that accounting rates of return are lower than in the no-growth case. Beaver and Ryan (2000) and Zhang (2001) model the interaction of conservatism and growth in a valuation context.

Our concern is not with the effects of conservatism and growth on accounting rates of return, but with the effects of the interaction between conservatism and *changes* in growth. Conservatism and growth do not necessarily affect the quality of earnings as we define the term. For example, if one expects a firm to maintain growth in net assets at the current level, then one expects future rates of return, *ceteris paribus*, to be the same as the current rate of return. However, should management reduce investment in assets that are subject to conservative accounting, the firm's rate of return will increase. If this reduction in investment is temporary, then so is the increase in the rate of return. In this case, the current rate of return will not be a good indicator of future rates of return.

We illustrate the effect on earnings of the interaction between conservative accounting and changes in investment with the case of LIFO accounting for inventories. The dollar effect of this interaction is transparent in the LIFO reserve disclosure in footnotes to U.S. financial statements. LIFO accounting is conservative because it carries inventories on the balance sheet at lower amounts than FIFO or at average cost methods if inventory prices

² The Gjesdal (1999) concept is subsumed within the Feltham and Ohlson (1995) and Beaver and Ryan (2000) concepts. However, the latter also include in the definition of conservative accounting the practice of recording positive net-present-value investments at unbiased historical cost to yield an accounting rate of return equal to the internal rate of return. Our characterization of conservatism does not involve this latter notion.

have risen in the past. Accordingly, accounting rates of return are typically higher under LIFO. The LIFO reserve is the difference between LIFO and FIFO carrying values.³ Correspondingly, the difference between LIFO and FIFO costs of goods sold is the change in the LIFO reserve. If purchases of inventories increase relative to sales (with continuing rising purchase prices), then earnings are lower under LIFO than in the no-growth case, and the LIFO reserve increases. Except for the case of the effect of extreme increases in purchase prices on cost of good sold, rates of return are also lower, for the denominator effect is typically greater than the earnings effect in the numerator. If dollar inventories decline (so that lower LIFO inventory costs flow into cost of goods sold), then LIFO yields higher earnings and rates of return. The LIFO reserve decreases, and earnings increase as the result of a real phenomenon: a decline in inventory. The buildup of the LIFO inventory creates the reserve, and the decline in inventory—known as LIFO dipping—is a liquidation of the reserve.

The same phenomena arise, although less transparently, from all forms of conservative accounting. Accelerated depreciation from using short, estimated asset lives (that reports lower net asset values) has no effect on earnings if depreciable assets are not growing, but reduces earnings (and, like LIFO, creates reserves) if investments in the assets increase, *ceteris paribus*.⁴ If investments decline, then accelerated depreciation creates earnings through the liquidation of reserves. Immediate expensing of R&D expenditures and advertising is conservative (setting knowledge assets and brand assets to zero on the balance sheet), but has no effect on earnings relative to capitalizing and amortizing the expenditures if the expenditures are not growing. However, increasing R&D investments and advertising, while immediately expensing the related costs, depresses earnings and increases reserves. Slowing R&D and advertising increases earnings and reduces reserves because the benefits from past expenditures are recognized without matching additional new expenditures.⁵ Because the reserves created by conservative accounting are not recorded on the balance sheet, they are unrecorded “hidden” reserves that are released into earnings when investment growth slows. We refer to them as “unrecorded reserves.”⁶

If a temporary change in investment leads to a temporary change in a reported accounting rate of return, then the current earnings will be a poor indicator of future earnings and, thus, earnings quality will be low. If an analyst accepts the current book rate of return as an unbiased indicator of future rates of return, then he will be misled. If investors value firms using current earnings as an input without appreciating that those earnings are unsustainable, then market valuations will also be of low quality. But, if the analyst and investor penetrate the joint effect of conservative accounting and investment, then they will discover that reported earnings is a poor predictor of long-run “sustainable” profitability, and will value the firm appropriately.

Our empirical analysis, then, documents the incidence of temporary effects of the interaction of conservative accounting and investment on accounting rates of return, and asks whether the market pricing of stocks is consistent with investors’ appreciating the quality of reported earnings in valuing stocks.

³ The SEC requires firms to calculate the LIFO reserve as the accumulated excess of the current cost of inventories over LIFO cost. With rapid inventory turnover, however, FIFO cost approximates current cost.

⁴ Conservative accounting reports lower earnings only in the growth case. Penman (2001, Chapter 17) gives examples of the effects of conservative accounting and growth described here.

⁵ Baber et al. (1991) and Perry and Grinaker (1994) provide evidence that managers modify R&D expenditures to manage earnings.

⁶ The LIFO reserve is, of course, recorded in the footnotes of the 10-K, but not in the balance sheet.

III. INDEXES OF CONSERVATISM AND EARNINGS QUALITY

We use information from the financial statements to develop two indices of the quality of earnings. The first index scores the degree to which the firm applies conservative accounting. The second index scores the quality of earnings that results from the joint effect of conservatism and changes investment activity.

Conservatism Index (C-Score)

The C-score measures the effect of conservative accounting on the balance sheet. We measure the C-score as the level of estimated reserves created by the conservatism, ER, relative to net operating assets:

$$C_{it} = \frac{ER_{it}}{NOA_{it}},$$

where i indicates firms and t indicates balance sheet dates.⁷ Net operating assets, NOA, is the book value of operating assets minus operating liabilities. That is, NOA equals assets minus liabilities, excluding financial assets and liabilities (financing debt and assets in which the firm invests excess cash). Balance sheets typically value these financial items at market value, or close to market value if interest rates change little, so conservative accounting does not affect such financial items.⁸ We subtract operating liabilities from operating assets in the NOA calculation to measure the net investment in operations. This netting recognizes that conservative accounting can affect both operating assets and operating liabilities.

A complete C-score calculates unrecorded reserves created by all operating items in the balance sheet—including bad-debt allowances, depreciation allowances, valuation allowances, deferred revenue, pension liabilities, and other estimated liabilities. We wish to distinguish earnings quality effects arising from changes in estimates (i.e., accounting manipulation) from quality effects arising from permanent accounting policy and changes in investment. An allowance for bad debts, for example, might be “high” because of a permanent policy of carrying net receivables at a conservative level or because of a temporary increase in the estimate of bad debts to reduce current income and increase future income. Because we cannot distinguish between the two effects for balance sheet items that are also subject to management’s estimates, we construct a C-score based on the accounting treatment of three investments for which the accounting is relatively immune from managerial discretion after the expenditure has occurred: inventories, R&D, and advertising. The accounting for these items follows mandates from accounting regulators or (in the case of LIFO) an accounting choice that management (usually) does not change from period to period. Thus:

⁷ One might consider developing a C-score calculated as the estimated reserve relative to net operating assets plus the estimated reserve—then the estimated reserve would be expressed as a percentage of net operating assets that would have been reported had the firm not practiced conservative accounting. In this paper, however, we wish to compare the effect of changes in reserves on return on net operating assets, and this return is, of course, denominated in net operating assets.

⁸ In terms of Compustat item numbers, net operating assets is calculated as Common Equity (60: Common Equity + 227: Preferred Treasury Stock - 242: Preferred Dividends in Arrears) + Financial Obligations (34: Debt in Current Liabilities + 9: Total Long-Term Debt + 130: Preferred Stock - 227: Preferred Treasury Stock + 242: Preferred Dividends in Arrears) - Financial Assets (1: Cash and Short-Term Investments + 32: Investments and Advances - Other) + Minority Interest (38: Minority Interest on Balance Sheet). The procedures for reformulating financial statements to separate operating and financing activities are laid out in Penman (2001, Chapters 8, 9, and 10).

$$C_{it} = (INV_{it}^{res} + RD_{it}^{res} + ADV_{it}^{res})/NOA_{it},$$

where the calculation for the three reserve components is as follows:

- Inventory reserve (INV_{it}^{res}) equals the LIFO reserve reported in the financial statement footnotes.
- R&D reserve (RD_{it}^{res}) is calculated as the estimated amortized R&D assets that would have been on the balance sheet if R&D had not been expensed. We capitalize R&D expenditures, then amortize them using the industry coefficients estimated by Lev and Sougiannis (1996). In a sensitivity analysis, we also amortize using the sum-of-the-years-digits method over five years.
- Advertising reserve (ADV_{it}^{res}) is the estimated brand assets created by advertising expenditures. We capitalize advertising expenses and then amortize them using a sum-of-the-year's digits method over two years. Bublitz and Ettredge (1989) and Hall (1993) indicate that advertising has a short useful life, typically one to two years.

We also calculate a subscore for each component of the reserve, the estimated reserve from that component relative to NOA.

Earnings Quality Indicator (Q-Score)

Whereas the C-score measures the effect of conservative accounting on the balance sheet, the Q-score measures the effect of conservative accounting on earnings in the income statement. Our Q-score is a combination of two measures. The first measure, Q_{it}^A , is calculated as:

$$Q_{it}^A = \frac{ER_{it}}{NOA_{it}} - \frac{ER_{it-1}}{NOA_{it-1}}$$

That is, Q_{it}^A is the change in the firm's conservatism score, C_{it} . Thus, a firm's Q^A score is positive if it builds up its reserve at a rate faster than that of the growth in net operating assets, and negative if it builds its reserve at a slower rate. The second measure, Q_{it}^B , compares a firm's C-score to the median for its SIC two-digit industry code:

$$Q_{it}^B = \frac{ER_{it}}{NOA_{it}} - \text{Industry median} \left(\frac{ER_{it}}{NOA_{it}} \right).$$

The Q-score combines these two measures:

$$Q_{it} = (0.5 \times Q_{it}^A) + (0.5 \times Q_{it}^B)$$

The Q-score does not ordinarily rank firms on their earnings quality. Both "high" (positive) and "low" (negative) Q-scores can indicate that current earnings are of poor quality, and a Q-score of zero indicates good quality. We also calculate Q subscores from C-scores for inventories, R&D, and advertising.⁹

⁹ Both the C-score and Q-score might be adjusted for deferred tax effects of the alternative accounting. This is not important for our purposes; Q is measured before tax. We use the score only to assign firms to high- and low-Q groups.

The change in estimated reserves measures the effect of conservative accounting and changes in investment on current earnings. To analyze the quality of the earnings, we focus on whether this effect is temporary. If a firm increased its reserve by growing investments but continued that growth in the future, then the effect would not be temporary. The Q^A score is the change in estimated reserves divided by net operating assets. Thus, the score is determined by the growth rate in the estimated reserves relative to the growth rate in net operating assets (so that, if estimated reserves grow at the same rate as net operating assets, the score is 0). Therefore, we benchmark a change in a firm's reserve against the change in the firm's net operating assets in the Q^A measure. Accordingly, we posit that the growth in the reserve (unrecorded assets) is unusual if it differs from the growth in (recorded) net operating assets. With the Q^B measure, we posit that if a firm's buildup of reserves relative to net operating assets—its C-score—differs from the industry median, then the firm's C-score will likely revert toward the industry median in the future. That is, the portion of the firm's C-score that deviates from the industry median is likely to be temporary. If, for example, a firm with a high C-score subsequently reverts toward the median, then, upon this reversion, some of the firm's estimated reserves will flow into earnings. Because it is not clear *ex ante* which benchmark (the firm's prior-period conservatism score or the current-period industry median score) might best capture temporary earnings effects, we weight the two measures equally, and then, in sensitivity analysis, we change the weights from (0.5, 0.5) to (1, 0) and (0, 1), to investigate separately Q^A and Q^B . The overall Q-score, as weighted, is:

$$\begin{aligned} Q_{it} &= 0.5 (C_{it} - C_{it-1}) + 0.5 (C_{it} - \text{Industry median } C_{it}) \\ &= C_{it} - (0.5 C_{it-1} + 0.5 \text{ Industry median } C_{it}). \end{aligned}$$

The overall score can thus be interpreted as the "unexpected" C, where the expected C is a simple average of last period's C and the median C for the industry.

Of course, our analyses are joint tests of the effect of the conservatism-investment interaction on earnings quality and of the ability of our scoring measures to capture conservatism and earnings quality. Thus, evidence that Q-scores identify unsustainable earnings will also validate a useful diagnostic tool, as an aid to financial statement analysis.

IV. SAMPLE SELECTION AND DESCRIPTION

We calculated C-scores and Q-scores for NYSE and AMEX nonfinancial firms on the combined Compustat Annual Industrial and Research files (which include nonsurvivors) for 1975–1997. We obtained monthly stock returns from the 1997 CRSP files. Our sample period begins in 1975 because, prior to that year, accounting data needed to construct the indices were missing from Compustat for a significant number of firms.

We were able to calculate net operating assets for 46,854 of the 46,955 firm-years for nonfinancial NYSE and AMEX firms with share price, shares outstanding, and book value of common equity data on Compustat at some time from 1975 to 1997. Of these, 46,122 firm-years had positive NOA, and there were sufficient data to compute C-scores for 38,540 firm-years. We deleted firm-years only if we could compute none of the three subscores from the Compustat data. When just one or two subscores were missing, we used industry median subscores (for two-digit SIC industry groups) as substitutes to calculate the overall C-score for that firm-year. The 38,540 firm-years included the following:

Firm-years with one subscore	8,285
Firm-years with two subscores	19,258
Firm-years with three subscores	<u>10,997</u>
Total firm-years	<u>38,540</u>

Of these 38,540 firm-years, the LIFO reserve was available for 36,244, and 13,931 reported using LIFO; therefore, there were 22,313 non-LIFO cases (where the reserve is 0). We could calculate the R&D reserve for 25,357 firm-years, and the advertising reserve for 18,191 firm-years.

Table 1 summarizes the distribution of C-scores over firm-years. The median of 0.114 indicates that NOA would have been 11.4 percent higher for the typical firm if the accounting treatment of the three items had not created reserves. The mean of 0.313, relative to the median and upper percentile scores, indicates that C-scores are particularly large for a relatively small number of firms (i.e., they are positively skewed).

We were able to compute Q-scores for 32,343 firm-years, but we deleted 2,547 firm-years for firms with non-LIFO inventory methods and missing C subscores for both R&D and advertising. Table 1 gives the distribution of the final sample of 29,796 Q-scores. The median is close to 0. The scores for the upper percentiles indicate that a nontrivial percentage of firms noticeably increased reserves during this period of considerable asset growth. Low Q-scores are smaller in absolute magnitude than high scores, but 40 percent of the firm-years reported liquidations of reserves.

TABLE 1
Distribution of Conservatism Index (C-Score) and Earning Quality Index (Q-Score) over Firm-Years; 1975–1997^a

	<u>C-Score</u>	<u>Q-Score</u>
Number of Firm-Years	38,540	29,796
Mean	0.313	0.099
Percentiles		
95	0.576	0.219
90	0.413	0.139
75	0.236	0.059
60	0.153	0.025
Median	0.114	0.009
40	0.084	0.000
25	0.048	–0.010
10	0.014	–0.046
5	0.004	–0.075

^a Section III of the text describes C-score and Q-score calculations. The C-score estimates the degree of understatement of net operating assets from the conservative practice of LIFO accounting for inventories and the expensing of R&D and advertising expenditures. It is measured as the unrecorded reserve from conservative accounting relative to recorded net operating assets.

The Q-score indicates the extent to which earnings are affected by change of investment under conservative accounting. Positive Q-scores indicate the extent to which earnings are temporarily depressed by growth in investment in inventories, R&D, and advertising. Negative Q-scores indicate the extent to which earnings are temporarily inflated by a decline of investment in these three items.

The mean first-order serial correlation of Q_{it}^A in the sample is 0.01, and that for Q_{it}^B is 0.61. In the cross-section, the mean Spearman rank correlation between Q_{it}^A and Q_{it+1}^A (in the subsequent year) is 0.10, decaying to 0.01 five years ahead. The mean Spearman rank correlation of Q_{it}^B with Q_{it+1}^B (in the subsequent year) is 0.88, decaying to 0.63 five years ahead. These statistics indicate that neither the Q^A nor the Q^B scores are permanent firm characteristics. However, they also indicate that, although the Q^B score does eventually revert toward the industry median, the Q^A score is very much a “one-time” feature, and so is more likely to capture the temporary effect on earnings that we wish to identify.

V. ANALYSIS OF THE QUALITY OF BOOK RATES OF RETURN

If the Q-score identifies temporary effects on earnings, then a relatively high Q-score should indicate current profitability that is lower than that expected in the future, and a relatively low Q-score should indicate current profitability that is higher than that expected in the future. We examine how Q-scores predict changes in core return on net operating assets from its current level. We define core return on net operating assets (RNOA) as:

$$\text{Core RNOA}_{it} = [\text{Core operating income}_{it} \times (1 - \text{statutory tax rate}_{it})] / \text{average NOA}_{it}.$$

Core operating income is operating income before interest, special items, extraordinary items, and discontinued operations. Core operating income therefore excludes transparently temporary components of earnings, along with interest expense (which is not generated by operations and which is unaffected by conservative accounting). Thus, we focus on reported income that investors might identify as sustainable income from operations. We measure after-tax core operating income without allocating taxes between operating and financing activities. Our measure assumes that any difference between effective tax rates on operating income and statutory rates (calculated as the Federal rate plus 2 percent) is temporary.

To conduct the analysis, we sorted firms in each year, 1976–1996 and within each two-digit SIC industry, into ten equal-sized groups based on the core RNOA the firm reported that year. We retained only RNOA groups with at least three firms. Then, within each RNOA group, we further divided firms into three equal-sized groups based on their Q-scores. We refer to these as high, median, and low Q-score groups. Our analysis compares core RNOA in the grouping year with core RNOA realized in subsequent years for each Q group. The grouping by industry controls for the type of operations, so the differences we observe in RNOA are more likely to be due to accounting effects than to differences in operations. The grouping also controls for the mean reversion in RNOA documented in Nissim and Penman (2001), so the observed relation between changes in RNOA over time and the Q-scores does not reflect this typical RNOA mean reversion.

Table 2 summarizes median core RNOA for the five years before and after the year of the grouping (Year 0) for each of these Q-score groups and for all firms together. The numbers in the table are means of the medians computed over the 21 calendar years from 1976 to 1996. Panel A of Figure 1 tracks these means for the high- and low-Q-score groups. Panel B tracks these same mean RNOAs after subtracting the median RNOA for all firms in that year to adjust for trends in median RNOA over time.

The results in Table 2 and Figure 1 indicate that, on average, Q-scores do discriminate on future core RNOA. The RNOA for high-Q groups are declining prior to Year 0, consistent with conservative accounting's increasing reserves and depressing earnings. The mean RNOA for these high-Q groups continues at about the same level, or slightly higher, after Year 0. The low-Q groups report lower mean RNOA than the high groups before Year 0, but by Year 0 their mean RNOAs are approximately the same as the high-Q groups'

TABLE 2
Summary of Median Core Return on Net Operating Assets (Core RNOA) in Years Relative to the Q-Scoring Year (Year 0), for Groups of Firms with High, Medium, and Low Q-Scores 1976–1996^a

	Year Relative to Year Q-Score is Calculated (Year 0)										
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Core RNOA											
High Q	0.1231	0.1214	0.1179	0.1143	0.1116	0.1107	0.1147	0.1122	0.1090	0.1088	0.1065
Medium Q	0.1165	0.1153	0.1136	0.1116	0.1097	0.1090	0.1075	0.1054	0.1026	0.1005	0.1013
Low Q	0.1113	0.1108	0.1094	0.1092	0.1110	0.1094	0.1002	0.0976	0.0950	0.0939	0.0930
All firms	0.1173	0.1159	0.1135	0.1117	0.1107	0.1096	0.1077	0.1052	0.1023	0.1012	0.1006
Differences in Core RNOA from Year 0											
High Q	0.0124	0.0106	0.0072	0.0036	0.0008	0	0.0039	0.0015	-0.0018	-0.0019	-0.0043
Medium Q	0.0075	0.0064	0.0046	0.0026	0.0008	0	-0.0014	-0.0036	-0.0063	-0.0085	-0.0076
Low Q	0.0020	0.0014	0.0000	-0.0002	0.0016	0	-0.0092	-0.0018	-0.0014	-0.0155	-0.0164
t-statistic on changes in Core RNOA from Year 0											
High Q	1.47	1.33	1.01	0.59	0.21	0	0.96	0.32	-0.07	-0.04	-0.38
Medium Q	0.97	0.91	0.74	0.47	0.22	0	-0.41	-0.61	-0.76	-0.90	-0.80
Low Q	0.23	0.20	0.00	-0.04	0.44	0	-2.78	-2.24	-2.30	-2.11	-2.01
Differences in changes in Core RNOA from Year 0 between High Q and Low Q groups	0.0104	0.0092	0.0071	0.0038	-0.001	0	0.0131	0.0132	0.0126	0.0136	0.0121
t-statistic on mean differences in changes in Core RNOA	5.650	4.324	4.092	1.950	-0.051		7.379	7.172	6.477	5.256	3.835

(Continued on next page)

TABLE 2 (Continued)

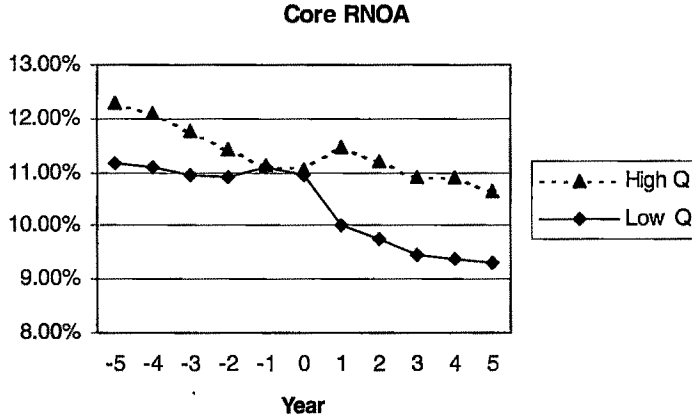
	Year Relative to Year Q-Score is Calculated (Year 0)										
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Number of Firms											
High Q	5,168	5,291	5,419	5,574	5,719	5,925	5,595	5,087	4,679	4,258	3,876
Medium Q	6,153	6,353	6,527	6,697	6,939	7,178	6,816	6,197	5,711	5,198	4,731
Low Q	4,184	4,324	4,486	4,683	4,897	5,193	4,928	4,475	4,094	3,728	3,398
All firms	15,505	15,968	16,432	16,954	17,555	18,296	17,339	15,759	14,484	13,184	12,005

^a The Q-score groups are based on a ranking of firms each year on Q-scores, within two-digit industry codes and within deciles of core RNOA. The high Q-score group is the top third of firms in that industry-core-RNOA ranking, the medium Q-score group is the middle third, and the low Q-score group is the bottom third of firms in the ranking on Q-score.

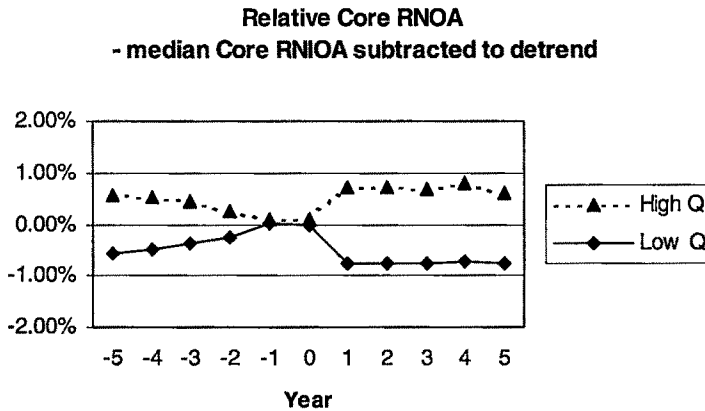
Core return on net operating assets (RNOA) is after-tax core operating income (exclusive of interest and unusual items) divided by average net operating assets. The core RNOA values reported in the table are the means of 21 yearly median core RNOAs computed over the years 1976 to 1996.

FIGURE 1
Mean Core Return on Net Operating Assets (Core RNOA) for High and Low Q-Score Groups
over Five Years before and after the Q-Scoring Year, Year 0

Panel A: Tracks Core RNOA



Panel B: Tracks Core RNOA Adjusted for the Median Core RNOA of All Firms in that Year



mean RNOAs (by construction of the tests). However, the mean RNOA for the low-Q groups subsequently deteriorates. Accordingly, the Q-score identifies, *ex ante* in Year 0, firms with low-quality earnings whose core RNOAs on average decline by the next year. By liquidating their reserves, these firms achieve an average core RNOA in Year 0 that is indistinguishable from the average RNOAs of the high-Q group and of all firms as a whole. The Q-score indicates, however, that this creation of earnings is temporary.

Table 2 reports two sets of significance tests. For the first set, we calculated median differences in core RNOA between Year 0 and each of the five years before and after Year 0 for high-Q, medium-Q, and low-Q groups, for each of the 21 years. The table reports the mean of these median differences over the 21 years, along with a t-statistic on those

mean differences (based on a standard error of the mean estimated from the time series of median differences). The second set of t-tests is on means of differences across high- vs. low-Q groups' differences in core RNOA (computed between Year 0 and the relevant year). The two sets of t-statistics indicate that the mean RNOAs for the low-Q groups in the five years after Year 0 are statistically significantly less than that in Year 0, and that the drop in core RNOA after Year 0 is statistically greater for the low-Q group than for the high-Q group. These differences—more than 1 percent of net operating assets—are economically significant.¹⁰ Of the low-Q firms that survived through Year +1, untabulated results reveal that only 44.2 percent had increasing RNOA in Year +1, compared to 52.5 percent for high-Q firms.¹¹ We repeated the analysis in Table 2 for the subperiods 1976–1982, 1983–1989, and 1990–1996, with similar results in each period.¹²

Sensitivity Tests and Further Analysis

We performed further tests to provide additional controls, to document the robustness of the findings over time and over Q subscores, and to explore the phenomenon further. We summarize these tests below; more detail is available on request.

Control for Mean Reversion in Core RNOA

We have attempted to control for the typical mean reversion of RNOA by comparing the evolution of core RNOA across Q groups for firms first grouped on the level of core RNOA. However, some of the differential behavior of core RNOA across Q-score groups might be attributable to differential behavior of different levels of RNOA if Q-scores are strongly correlated with RNOA (and thus the ranking on Q-scores is effectively a ranking on RNOA). Untabulated analyses suggest this is not a problem. The median Spearman correlation between Q-scores and RNOA within RNOA groups is 0.00, the mean correlation is 0.10, and the 75th percentile of rank correlation is 0.42.¹³ We also repeated the analysis in Table 2, first with a ranking on core RNOA within each RNOA group, then splitting into three RNOA groups rather than on Q-scores. In contrast to the results reported in Figure 1 and Table 2, which are based on Q-score partitions, we observed little difference in the RNOA dynamics across these three RNOA groups, so mean reversion in RNOA cannot explain the results.

Analysis of Inventory, R&D, and Advertising Subscores

We performed an analysis identical to that in Table 2, except that we used separate Q-scores for inventory, R&D, and advertising. Results (not tabulated here) are similar for each subscore, so the results are not due to one particular accounting item.

We next estimated 21 annual linear regressions of one-year-ahead core RNOA (in $t + 1$) on core RNOA and Q-scores in the scoring year, t . Panel A of Table 3 shows that the positive mean estimated coefficient for the Q-score and its t-statistic indicate that the

¹⁰ Using the residual income valuation formula, a revision of forecasted RNOA of 1.3 percent per year for five years has considerable effect on the calculated value.

¹¹ For a classification of high-Q and low-Q firms into those with increasing and decreasing core RNOA in Year 1, the Chi-square statistic is 73.9, with a probability given no relationship between Q and subsequent profitability of less than 0.001.

¹² Although firms in the same industry tend to have the same fiscal year-ends, they do vary. Thus, ranking firms as if their RNOAs were available at the same time is subject to a peeking-ahead bias. However, we repeated the analysis on firms with (the same) December 31 fiscal year-ends, with similar results.

¹³ The mean annual Spearman rank correlations between the Q-score and core RNOA over all firms was 0.08, and the mean for ranking within industry was 0.07. Thus, in addition to the low correlation between Q-score and core RNOA within RNOA groups, correlation between the Q-score and core RNOA is minimal even over all levels of RNOA.

TABLE 3
Results of Regression Tests of the Ability of Q-Scores to Predict Year-Ahead Core RNOA^a

Panel A: Overall Q-Score

$$RNOA_{t+1} = \alpha_0 + \alpha_1 RNOA_t + \alpha_2 Q + e_{t+1}$$

	<u>Mean</u>	<u>First Quartile</u>	<u>Median</u>	<u>Third Quartile</u>
Intercept	0.016 (6.51**)	0.009	0.018	0.022
RNOA coefficient	0.800 (47.55**)	0.782	0.813	0.846
Q coefficient	0.096 (7.17**)	0.054	0.103	0.136

Panel B: Three Subscores for Inventory, Research and Development, and Advertising

$$RNOA_{t+1} = \alpha_0 + \alpha_1 RNOA_t + \alpha_2 Q\_inv + \alpha_3 Q\_rd + \alpha_4 Q\_adv + e_{t+1}$$

	<u>Mean</u>	<u>First Quartile</u>	<u>Median</u>	<u>Third Quartile</u>
Intercept	0.015 (6.39**)	0.010	0.016	0.022
RNOA coefficient	0.802 (43.56**)	0.765	0.823	0.866
Q_inv coefficient	0.131 (6.18**)	0.076	0.133	0.215
Q_rd coefficient	0.106 (5.15**)	0.080	0.130	0.149
Q_adv coefficient	0.082 (2.49*)	0.007	0.071	0.195

*, ** Denotes significance at the 0.05 and 0.01 levels, respectively, using a two-tailed t-test.

^a We estimated regression coefficients for 21 annual cross-sectional regressions, 1976 to 1996. The mean estimated coefficients from the 21 regressions appear in the table, along with those at the 25th percentile, the median, and the 75th percentile. The t-statistics (in parentheses) are calculated as the mean of the 21 estimated coefficients relative to their estimated standard errors.

RNOA is core return on net operating assets; Q is the Q-score that indicates temporary effects of conservatism and changes in investment on current earnings; Q_inv is the inventory subscore; Q_rd is the research and development subscore; Q_adv is the advertising subscore.

Q-score improves forecasts of one-year-ahead RNOA over that from the current RNOA alone. Panel B reports the results of a similar analysis, after separating the Q-score into its three subscores. The results indicate that each subscore provides additional information about one-year-ahead RNOA, not only beyond current RNOA, but also relative to each other.¹⁴

¹⁴ The mean Spearman rank correlation between the score for inventory (Q_inv) and those for R&D (Q_rd) and advertising (Q_adv), within industry, were 0.09 and 0.08, respectively. The correlation between Q_rd and Q_adv was 0.11.

Analysis of Q^A and Q^B

The analyses in Tables 2 and 3 are based on a Q-score that weights Q^A and Q^B equally. As explained in Section III, we intended both of these measures to identify temporary changes in reserves generated by conservative accounting. Additional tests reveal that both measures, Q^A and Q^B , are effective in forecasting future core RNOA. However, Q^A appears to be somewhat more effective than Q^B , consistent with our earlier conjecture that the lower serial correlation in Q^A than in Q^B indicated that Q^A better captures the temporary effect of the conservative accounting on earnings.

Specifically, we repeated the tests in Table 2 for both Q^A and Q^B . Although Q^A produced slightly stronger results than Q^B , both discriminate on changes in future core RNOA. The mean changes in core RNOA from Year 0 to Years +1 to +5 for the low- Q^A group were -0.95 percent, -1.03 percent, -1.19 percent, -1.16 percent, and -1.11 percent, respectively. For the low- Q^B group the mean changes for these five years were -0.68 percent, -0.99 percent, -1.42 percent, -1.49 percent, and -1.70 percent, respectively. All the t-statistics were of the same order as those reported in Table 2. For the high-Q groups, results were stronger for the Q^A metric, with mean core RNOA increasing by 0.65 percent in Year +1 with a t-statistic of 1.78, significant at the 0.05 level for a one-tailed test. For Q^B , mean core RNOA increased by 0.26 percent in Year +1 with a t-statistic of 0.65.

We also repeated the regression tests reported in Table 3 for Q^A and Q^B . The results are similar. The mean estimated coefficient for Q^A corresponding to that in Panel A of Table 3 was 0.212 (with a t-statistic of 6.34) and that for Q^B was 0.030 (with a t-statistic of 4.70). Mean coefficient estimates on RNOA were similar to those in Table 3. For both Q-scores, coefficients on all three subscores (corresponding to those in Panel B of Table 3) were significantly positive. Again, however, Q^A measures provided more explanatory power than Q^B measures.

Control for Growth in Net Operating Assets

The Q^A -score captures the rate of change in estimated reserves relative to the rate of change in net operating assets. By benchmarking growth in estimated reserves (from unrecorded assets) against growth in recorded assets, Q^A is designed to capture abnormal growth in estimated reserves. However, the Q^A growth measure is clearly affected by the growth in net operating assets.

Fairfield et al. (2001) show that growth in net operating assets reduces future profitability, and they argue that Sloan's (1996) observation that high (low) accruals are followed by lower (higher) profitability is due to the positive correlation between growth and accruals. The negative effect of growth of net operating assets on future profitability can be attributed to new investment earning at a lower rate or to the effect of conservative accounting; new investment, amortized rapidly, reduces subsequent earnings. Although the latter explanation involves conservative accounting, we have a different feature of conservative accounting in mind.

Therefore, to ensure that our inferences are not solely attributable to differential growth in net operating assets, we estimated the cross-sectional regressions in Panel A of Table 3, adding the rate of growth in net operating assets as an additional explanatory variable. Growth in net operating assets indeed plays an incremental role in explaining one-year-ahead core RNOA (with a statistically significant negative mean estimated coefficient), but the Q-score still provides incremental information, with a positive mean coefficient of 0.039 and a significant t-statistic of 3.23.

Further Analysis of the RNOA Changes

Our analysis investigates whether core RNOA is temporarily affected by increases in unrecorded reserves that depress earnings, and by the release of those reserves that increase earnings. Core RNOA equals core profit margin (core operating income/sales) multiplied by asset turnover (sales/net operating assets), following the standard DuPont decomposition. To confirm that the effect was indeed an effect on earnings (and thus the profit margin) and not just a turnover effect, we examined the two components separately.

Table 4 reports core profit margins and asset turnovers for each of the five years before and after the Q-scoring year, for both high- and low-Q groups. As the accounting for inventories, R&D, and advertising does not affect depreciation and amortization, profit margins are before depreciation and amortization.

The average changes in profit margins from Year 0 to subsequent years are significantly different for the low-Q groups than for the high-Q groups. Profit margins increase for the high-Q groups, and decrease for low-Q groups.¹⁵ We also performed this analysis for core profit margins before R&D, and advertising expense (and depreciation and amortization). For the low-Q group, these margins declined in Year 0 on average, whereas those after R&D and advertising expenses increased. Thus, lower R&D and advertising expenditures turned otherwise decreasing mean margins into increasing margins.

Table 4 shows that mean asset turnovers increase after Year 0 for high-Q firms but decline for low-Q firms; thus, the Q-score predicts differences in changes in asset turnovers, as well as margins. Note, furthermore, that because the mean asset turnover for low-Q groups declines in Year 0 (reducing RNOA), the core RNOA created by low-Q firms in Year 0 appears to be a profit margin effect. The decline in asset turnovers for low-Q firms in Year 0 is driven by sales' increasing at a slower rate than net operating assets. Thus, the low-Q firms are increasing profit margins and maintaining core RNOA in Year 0 at the level of high-Q firms, on average, even though their sales growth is declining relative to the growth of their net operating assets. Furthermore, the decline in asset turnover subsequently continues, on average. Asset turnovers for high-Q groups increase in Year 0. These results indicate that changes in turnover are related to earnings quality, as in Fairfield and Yohn (2001) but for different reasons, for that paper views changes in net operating assets (rather than Q) as a diagnostic of earnings quality.

Analysis of Changes in Investment

To corroborate that changes in Q are indeed due to changes in investment, we tracked the investment in inventory, R&D, and advertising. For these three items, changes in reserves due to conservative accounting result from changes in dollar investments, so we tracked cumulative (compounded) growth rates of estimated reserves for high- and low-Q groups for years -5 to +5. The average growth rates for the low-Q groups declined in years up to and including Year 0, and were lower than those for high-Q firms. However, they subsequently converged toward the high-Q growth rates indicating that the slowing of investment growth up to Year 0 was indeed temporary, and thus increased earnings, profit margins, and RNOA temporarily.

There is one qualification to our interpretation of the results. Firms that reduce investment or the rate of growth in investment can do so in anticipation of lower profitability of new projects. Therefore, the lower RNOA after Year 0 for low-Q firms may reflect lower

¹⁵ Regression tests similar to those in Table 3 support the inferences from Table 4. Also, of the firms that survived through Year +1, 49.8 percent of low-Q firms had increasing profit margins in Year +1, compared to 55.1 percent of high-Q firms.

TABLE 4
Summary of Core Profit Margins (before Depreciation and Amortization) and Asset Turnovers in Years Relative to the Q-Scoring Year (Year 0) for Groups of Firms with High, Medium, and Low Q-Scores 1976-96^a

	Year Relative to Year Q-Score is Calculated (Year 0)										
	-5	-4	-3	-2	-1	0	1	2	3	4	5
<i>Panel A: Core Profit Margin</i>											
High Q	0.1215	0.1198	0.1173	0.1147	0.1148	0.1139	0.1164	0.1169	0.1167	0.1174	0.1172
Medium Q	0.1201	0.1202	0.1214	0.1211	0.1212	0.1215	0.1215	0.1217	0.1218	0.1203	0.1200
Low Q	0.1154	0.1168	0.1164	0.1200	0.1200	0.1251	0.1225	0.1204	0.1210	0.1209	0.1200
All firms	0.1192	0.1192	0.1191	0.1187	0.1187	0.1196	0.1200	0.1194	0.1192	0.1187	0.1184
Differences in changes in core profit margins from Year 0: High Q - Low Q	0.0173	0.0143	0.0121	0.0060	0.0027	0	0.0051	0.0072	0.0059	0.0066	0.0067
t-statistic on mean differences in changes in core profit margins	8.01	6.61	7.53	4.19	2.25		3.82	5.29	4.40	4.45	4.07
<i>Panel B: Asset Turnover</i>											
High Q	2.222	2.245	2.272	2.279	2.294	2.361	2.405	2.381	2.355	2.322	2.290
Medium Q	2.123	2.116	2.119	2.105	2.101	2.111	2.093	2.094	2.086	2.065	2.070
Low Q	2.054	2.069	2.058	2.068	2.080	2.015	1.961	1.958	1.953	1.948	1.973
All firms	2.130	2.151	2.106	2.166	2.172	2.179	2.163	2.153	2.143	2.132	2.125
Differences in changes in asset turnovers from Year 0: High Q - Low Q	-0.178	-0.170	-0.132	-0.135	-0.132	0	0.0974	0.0856	0.0783	0.0592	0.0165
t-statistic on mean differences in changes in asset turnover	-5.92	-5.32	-5.12	-5.45	-6.49		5.25	4.00	3.05	2.22	0.68

^a The Q-score groups are based on a ranking of firms each year on Q-scores, within two-digit industry codes and deciles of core RNOA. The high Q-score group consists of the top third of firms in that industry-core-RNOA ranking, the medium Q-score group consists of the middle third, and the low Q-score group consists of the bottom third of firms on Q-scores. Core profit margin is core operating income before depreciation and amortization divided by sales. Asset turnover is sales divided by net operating assets. The core profit margin and asset turnover numbers in the table are the means over 21 yearly median numbers computed over the years 1976-1996.

project profitability. The finding that the decline in investment is temporary argues against this interpretation. Furthermore the conservatism effect is at work, *ceteris paribus*, by the construction of the accounting. In any case, an anticipated change in profitability from its current level for any reason indicates unsustainable current earnings, and the results indicate the Q-score identifies unsustainable earnings irrespective of the cause.

Survivorship

We have presented the core RNOA for high- and low-Q groups for Years +1 to +5 as those that an investor might predict by observing Q-scores in Year 0. However, the analysis is of *ex post* outcomes, and we included firms in the analysis for Years +1 to +5 only if they survived to the respective years. Therefore, we do not observe realizations of core RNOA for the nonsurvivors that would have been predicted *ex ante*. This is a problem only if survivorship differs over high- and low-Q groups.

The number of firms included in the analysis for each of these years appears at the bottom of Table 2.¹⁶ The survival rate for high-Q firms in Years +1 to +5 were 95.9 percent, 92.5 percent, 89.3 percent, 86.5 percent, and 83.8 percent, respectively—very similar to the corresponding rates for low-Q firms of 96.3 percent, 92.7 percent, 89.6 percent, 86.7 percent, and 83.8 percent, respectively. Of the high-Q firms that did not survive, 94.6 percent failed to survive because of merger and acquisition; the corresponding percentage for low-Q firms was only 82.5 percent. Thus, low-Q firms had a high percentage of firms that terminated for reasons other than mergers, such as liquidation, delisting, or suspension of trading.

VI. THE RETURNS TO QUALITY ANALYSIS

We now investigate whether the stock market prices stocks as if it appreciates the differential quality of earnings indicated by our Q-score. We take investment positions in stocks in the sample period based on their Q-scores and observe whether these positions earn differential returns, adjusting for conjectured risk differentials.

In each year from 1976 to 1995, and within each two-digit SIC industry code, we ranked the firms into ten equal-sized portfolios based on their Q-scores. Year 0 is the year we formed the portfolios. The portfolio formation date is three months after fiscal year-end, by which time the firm must file its annual reports with the SEC. We then calculated mean buy-and-hold returns for each portfolio for each year, -2 to +5, relative to this date. The computed returns include delisting returns for nonsurvivors. Table 5 reports mean raw returns and size-adjusted returns during the 20 years that the positions were taken. The grouping within industry controls for operating risk (to some degree), and the size adjustment controls for the stock return “size effect” that researchers (e.g., Fama and French 1992) conjecture is a premium for risk. We computed the size-adjusted returns by subtracting the raw (buy-and-hold) return on a size-matched, value-weighted portfolio formed from size-decile groupings supplied by CRSP.

The mean returns for years +1 to +5 in Table 5 are positively related to Q. The size-adjusted returns for portfolios 7–10 (the high-Q portfolios) are positive, and those for Year +1 (in boldface) are significantly different from 0. Those for the low-Q portfolios 1–4 are considerably lower, and many are negative. The difference between the mean raw returns

¹⁶ The number of firms at the bottom of Table 2 for Years +2 to +5 overstates survivorship. Because the data period ended in 1997, data were not available for all of the Years +2 to +5 for firms with Year 0 rankings in 1993–1996. The survivorship rates that follow exclude the observations unavailable at the end of the sample period.

TABLE 5
Mean Percentage Annual Stock Returns for Portfolios Formed on Q-Scores 1976–1995^a

<i>Q</i> Portfolios	<i>Year Relative to Year <i>Q</i>-Score Is Calculated (Year 0)</i>							
	<i>–2</i>	<i>–1</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
<i>Panel A: Raw Returns</i>								
Lowest Q	26.67	28.19	26.39	17.03	22.51	20.81	22.58	19.12
2	21.27	22.25	19.51	19.02	20.36	21.20	18.70	17.15
3	22.56	20.75	21.16	21.39	20.90	20.68	18.94	20.26
4	23.25	24.39	21.01	19.63	22.85	20.35	24.13	18.21
5	21.30	23.44	19.81	20.87	21.73	20.75	22.19	17.52
6	20.87	22.37	19.56	20.97	20.94	20.15	21.44	19.80
7	20.32	21.58	20.15	22.64	21.47	20.77	20.55	18.24
8	20.57	22.39	21.20	21.19	21.05	21.41	19.39	17.84
9	19.66	18.50	21.94	21.20	21.64	19.84	19.70	18.88
Highest Q	18.25	20.64	23.33	26.06	23.53	22.78	22.12	20.93
High – Low	–8.41	–7.55	–3.06	9.03	1.02	1.98	–0.46	1.80
Significance	0.000	0.000	0.013	0.000	0.265	0.091	0.616	0.136
<i>Panel B: Size-Adjusted Returns</i>								
Lowest Q	3.39	5.40	4.21	–3.17	–0.26	–0.09	–0.24	–0.06
2	0.08	0.72	–0.57	–0.66	0.48	1.58	–1.29	–0.22
3	2.33	0.72	1.21	1.72	1.35	1.05	–0.47	3.90
4	2.97	3.27	1.56	0.12	2.58	1.09	5.34	1.99
5	1.32	2.68	0.54	1.51	1.46	2.15	2.23	0.65
6	1.07	2.20	0.27	1.06	1.78	1.59	3.47	2.86
7	1.34	2.41	0.87	2.85	2.12	1.99	2.09	1.79
8	2.61	1.23	3.14	2.91	1.44	3.05	0.80	1.29
9	0.79	–0.12	2.10	2.76	2.40	1.02	0.39	2.06
Highest Q	–0.42	0.19	2.34	5.78	1.80	2.38	2.63	2.90
High – Low	–3.81	–5.21	–1.87	8.95	2.05	2.46	2.86	2.96
Significance	0.001	0.000	0.071	0.000	0.070	0.036	0.024	0.031

^a We calculated buy-and-hold returns for each Q-score portfolio for each year, –2 to +5, in each scoring year, 1976–1995. Panel A reports the mean returns over the 20 years. Panel B reports the mean size-adjusted returns, computed by subtracting the raw (buy-and-hold) return on a size-matched, value-weighted portfolio formed from CRSP size-deciles.

The high-low return is the return from investing long in the highest Q portfolio and investing the same dollar amount short in the lowest Q portfolios, for 0 net investment.

Significance tests are based on 5,000 replications of randomly assigning firms to Q portfolios. The significance levels for the high-low return are the frequency of observing returns equal to the positive (negative) return on the high-low investment strategy, or higher (lower), in the 5,000 replications. For the size-adjusted returns in Panel B, returns in boldface are those where we observed returns equal to or greater than the return reported (or equal to or less than the return reported if negative) less than 5 percent of times in 5,000 replications with random assignment of stocks to portfolios.

for the highest Q and lowest Q portfolios in Year +1 is 9.03 percent, or 8.95 percent on a size-adjusted basis. This is a return, before transaction costs, to a zero-net-investment strategy with the canceling of long and short positions in the highest- and lowest-Q portfolios. We assessed the statistical significance of these returns by randomly assigning firms to Q portfolios in 5,000 replications of the zero-net-investment strategy. The reported significance levels are the relative frequencies of observing the actual mean differences, or higher, in these replications. The results in Table 5 reveal that most of the return from going long on high-Q stocks and shorting low-Q stocks occurs in the first year after Year 0. Figure 2 indicates that this strategy would have yielded a positive return before transaction costs in every year except 1990.

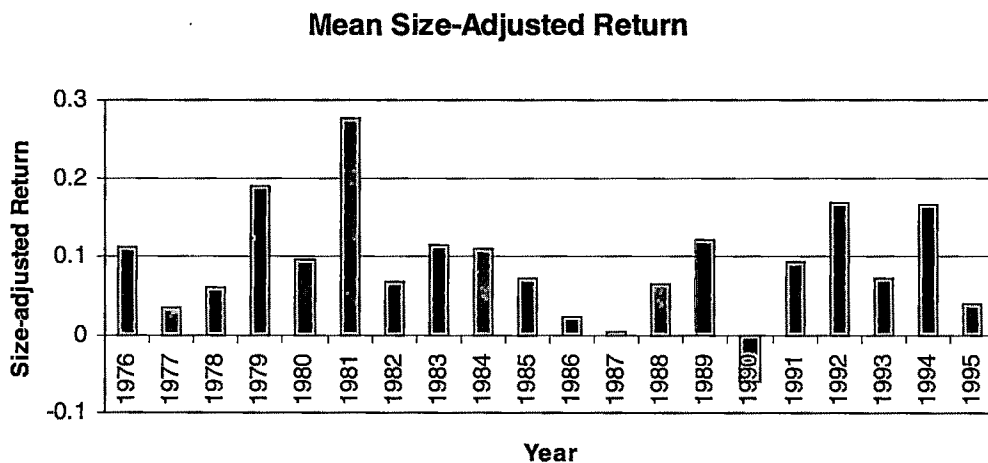
Not only are the returns to the zero-net-investment strategy positive after Year 0, but they are also negative prior to Year 0. The difference in returns between high and low groups in Years -2 to 0 and Years +1 to +5 correspond to differences in the changes in core RNOA graphed in Figure 1. Over Years -2 to 0, the average median-adjusted core RNOA increases for low-Q firms (releasing reserves), compared to the decrease for high-Q firms (building up reserves), with a reversal of the pattern in Years +1 to +5. So these return reversals are consistent with the market interpreting the firms' reported earnings uncritically in Years -2 to 0, but correcting this mistake in subsequent years. It appears that the market fails, at least in part, to penetrate the earnings quality due to changes in reserves created by conservative accounting and changes in investment.

Sensitivity Tests and Further Analysis

Implementation of the Trading Strategy

The trading strategy, as implemented in Table 5, may suffer from a "peeking-ahead" bias. We used Lev and Sougiannis's (1996) estimated coefficients to capitalize and amortize R&D, and these estimates use data from dates after we take positions in the stocks, at least for some years. Furthermore, in calculating the C-score, we substituted industry medians when certain subscores were missing, but median data are sometimes unavailable at the

FIGURE 2
Mean Size-Adjusted Return Differences between High-Q and Low-Q Portfolios in the Year Following the Q-Scoring (Year $t + 1$), for Each Year, 1976–1995



ranking date due to different fiscal year-ends. The Q^B component of the Q-score also involves an industry comparison. Finally, because firms do not have the same fiscal year-ends, their Q-scores would not have been available at the same time for ranking relative to each other.

Accordingly, we repeated the tests reported in Table 5 after modifying the research design to eliminate each of these possible biases. We capitalized R&D, using a sum-of-the-year's digits method over five years to amortize capitalized expenditures. We used the prior year's industry median for calculating C-scores, and, when forming portfolios, we compared Q-scores only for firms with December 31 fiscal year-ends. With these modifications, the difference in size-adjusted returns between high- and low-Q portfolios in Year +1 (with significance levels in parentheses) was 7.63 percent (0.000), and, for Years +2 to +5, the difference was 3.36 percent (0.018), 2.97 percent (0.020), 0.36 percent (0.516), and 3.92 percent (0.016), respectively, for the 58.8 percent of firm-years with December 31 fiscal year-ends.¹⁷

Dealing with Possible Late Filers

For the analysis in Table 5 we took positions in stocks three months after fiscal year-end, based on the assumption that annual reports are publicly available by that time. Although the SEC requires firms to file their financial reports within those three months, some firms file late. Table 6 reports monthly mean size-adjusted returns for each month beginning four months after fiscal year-end. The returns are for the low-Q and high-Q portfolios in Table 5, and for the difference between the two. It is clear from the returns for month 5 onward that the Year +1 returns in Table 5 are not due solely to returns in the fourth month after the fiscal year-end, when firms may still be reporting.

Table 6 also reports the number of quarterly earnings announcements in each month. The returns to long positions in high-Q firms and the canceling of short positions in the low-Q firms are relatively high in months 7 and 8 when second-quarter earnings for Year +1 would have been announced, and in months 10 and 11 when third-quarter earnings for that year would have been announced. This, like similar investigations of how apparent mispricing corrects itself (e.g., Freeman and Tse 1989; Bernard et al. 1997), suggests the market corrects its mispricing of forecasted earnings for Year +1 as those earnings are realized. The association of the returns with earnings reporting also supports the interpretation that the predicted returns are attributable to mispricing rather than to rewarding for risk. However, the association of the returns in Table 6 with earnings reporting months is imperfect; for months 4 and 5 when first quarter earnings would have been announced, the returns are relatively low, and returns in some months with relatively few announcements are positive and significant. Tests of returns around narrower windows surrounding subsequent earnings announcements and analysts' forecast revisions would be necessary to draw more definitive conclusions.

Analysis of Potential Risk Proxies

Risk is nearly always a potential alternative explanation for predictable abnormal returns. Most of the return differences between high-Q and low-Q groups occur in Year +1,

¹⁷ We also made a calculation for all firms irrespective of fiscal year-end, by forming portfolios every month based on firms with fiscal year-ends three months before, and then weighting the monthly portfolios equally (investing \$1 each month). The mean, size-adjusted return difference (significance level) between high- and low-Q portfolios in Year +1 was 15.27 percent (0.004) using equal weights, and 6.34 percent (0.072), 7.27 percent (0.056), -0.20 percent (0.460), and 5.79 percent (0.147) in Years +2 to +5, respectively. We obtained similar results by weighting monthly portfolios according to the number of firms in the month's ranking.

TABLE 6
Month-by-Month Size-Adjusted Returns after Fiscal Year-End for Portfolios Formed on Q-Scores, and the Number of Quarterly Earnings Announcements in Each Month^a

	Month after Fiscal Year-End											
	4	5	6	7	8	9	10	11	12	13	14	15
Low-Q portfolio return	1.07	-0.14	0.39	0.02	-0.23	-0.45	-0.39	-0.50	-0.52	0.33	0.54	0.09
High-Q portfolio return	1.16	0.09	0.03	0.94	0.50	0.05	0.71	0.49	0.38	1.33	0.91	0.28
High - Low return	0.09	0.23	-0.36	0.92	0.73	0.50	1.10	0.99	0.90	1.00	0.37	0.17
Number of quarterly earnings announcements	2,330	1,127	36	2,138	1,133	32	2,190	1,072	22	687	1,739	693

^a Returns are monthly size-adjusted returns for the lowest and highest Q portfolios in Table 5. The high - low return is the difference in the size-adjusted returns between the high and low Q portfolios. Q is the Q-score. Returns in boldface are significantly different from 0, with significance assessed by observing less than 5 percent of returns greater than the reported amount (or less than the reported amount, if negative) in 5,000 replications with random assignment of stocks to portfolios. The number of quarterly earnings announcements is the number of announcements made in the respective month after fiscal year-end.

suggesting a temporary correction to market inefficiency rather than a permanent effect that one would expect if the differences were due to risk. However, the return differences are positive in Years +2 to +5, and significantly so for the size-adjusted returns, so part of the observed returns to the Q-strategy could be attributable to risk. The specification and measurement of risk premiums are elusive. Our design, which matches on industry and size, mitigates differences in risk across Q-groups. Furthermore, in untabulated analyses we found that Q is not highly correlated with factors that Fama and French (1992) nominate as risk proxies: the mean, cross-sectional correlation between the Q-score and beta, book-to-market, leverage, and earnings-to-price are -0.00, -0.04, -0.09, and -0.02, respectively. Finally, in Table 7 we report the results of estimating Fama and Macbeth (1973) type cross-sectional regressions, using individual stocks, with a control for factors that Fama and French (1992) suggest are risk factors:

$$R_{i,t+1} = \alpha_0 + \alpha_1 \beta_{i,t} + \alpha_2 \ln(M)_{i,t} + \alpha_3 \ln(B/M)_{i,t} + \alpha_4 \ln(LEV)_{i,t} + \alpha_5 (E(+)/P)_{i,t} \\ + \alpha_6 (E/P \text{ dummy})_{i,t} + \alpha_7 Q_{i,t} + e_{i,t}$$

where:

- $R_{i,t+1}$ = annual return in Year +1 after the Q-scoring (the year begins three months after fiscal year-end);
- $\beta_{i,t}$ = Capital Asset Pricing Model beta;
- $M_{i,t}$ = market value;
- $(B/M)_{i,t}$ = book-to-market ratio;
- $(LEV)_{i,t}$ = leverage, calculated as book value of total assets to book value of equity;
- $(E(+)/P)_{i,t}$ = earnings-to-price ratio, positive earnings only;
- $(E/P \text{ dummy})_{i,t}$ = negative earnings dummy, 1 if earnings are negative, 0 otherwise; and
- $Q_{i,t}$ = Q score in year t.

Results are in Table 7 with and without including the Q-score in the regression. The mean estimated coefficient on Q is significantly positive. We conclude that Q-scores forecast returns in excess of those expected from commonly identified firm-risk factors.

Control for the Accrual Anomaly

One must be cautious in attributing abnormal returns to particular predictors; the correlation between returns and a particular predictor may arise because that predictor is correlated with another predictor. Sloan (1996) identified abnormal accruals as a predictor of abnormal returns. Accruals are a different accounting phenomenon than the conservative accounting effect investigated here, but Q-scores and accruals might be correlated if managers use accruals as well as real investment to manage earnings.

Accordingly, we estimated the regressions in Table 7 including an accrual variable, defined (as in Sloan 1996) as the change in noncash current assets, less the change in current liabilities exclusive of short-term debt and taxes payable, less depreciation expense, all divided by average total assets. The estimated coefficient on the accrual variable was -0.424, with a significant t-statistic of -4.39 (consistent with Sloan's [1996] analysis), and the coefficient on the Q-score was 0.221, with a t-statistic of 2.65. When we defined the accrual variable (more appropriately) as the difference between net income and cash from

TABLE 7
Mean Coefficient Estimates for Regressions of Year-Ahead Returns on Q-Scores and Factors to Control for Risk 1976–1995^a

$$R_{i,t+1} = \alpha_0 + \alpha_1 \beta_{i,t} + \alpha_2 \ln(M)_{i,t} + \alpha_3 \ln(B/M)_{i,t} + \alpha_4 \ln(LEV)_{i,t} \\ + \alpha_5 (E(+)/P)_{i,t} + \alpha_6 (E/P \text{ dummy})_{i,t} + \alpha_7 Q_{i,t} + e_{i,t}$$

Variable	Definition	Q-Score Included in the Regression		Q-Score Not Included in the Regression	
		Coefficients	t-statistics	Coefficients	t-statistics
Constant	Intercept	0.2601	5.680**	0.2740	6.017**
$\beta_{i,t}$	Beta	-0.0104	-0.573	-0.0009	-0.505
$\ln(M)_{i,t}$	Size	-0.0192	-3.333**	-0.0185	-3.213**
$\ln(B/M)_{i,t}$	Book-to-market	0.0276	1.933	0.0251	1.724
$\ln(LEV)_{i,t}$	Leverage	0.0491	2.634*	0.0430	2.227*
$(E(+)/P)_{i,t}$	E/P	0.2878	1.435	0.2017	1.033
$(E/P \text{ dummy})_{i,t}$	Negative E dummy	-0.0109	-0.322	-0.0083	-1.517
$Q_{i,t}$	Q-score	0.2873	3.834**		

*, ** Denotes significance at the 0.05 and 0.01 levels, respectively, using a two-tailed test.

^a Regression variables are as follows:

$R_{i,t+1}$ = annual return in Year +1 after the Q scoring; the year begins three months after fiscal year-end;

$\beta_{i,t}$ = Capital Asset Pricing Model beta;

$M_{i,t}$ = market value;

$(B/M)_{i,t}$ = book-to-market ratio;

$(LEV)_{i,t}$ = leverage, calculated as book value of total assets to book value of equity;

$(E(+)/P)_{i,t}$ = earnings-to-price ratio, positive earnings only;

$(E/P \text{ dummy})_{i,t}$ = negative earnings dummy, 1 if earnings are negative, 0 otherwise; and

$Q_{i,t}$ = Q score in year t.

We estimated regression coefficients for each month of the sample period, where firms in the estimation for each month are those whose fiscal years ended three months earlier. Coefficients reported are mean regression estimates over 240 monthly cross-sectional regressions from 1976 to 1995.

operations (divided by average total assets), the estimated coefficient on accruals was -0.302, with a t-statistic of -3.46, and that on the Q-score was 0.260, with a t-statistic of 3.12. We therefore conclude that the effect of the Q-score is incremental to that of accruals relative to total assets.

Analysis of Subscores

We repeated the return tests substituting Q-scores calculated from C subscores: (1) inventory accounting, (2) R&D, and (3) advertising. Substituting each subscore in separate regressions yielded results quite similar to those for the composite score. The mean, size-adjusted return for the zero-net-investment strategy in Year +1 for the R&D subscore was 7.33 percent (significance level 0.000), 5.22 percent (0.000) for the inventory subscore, and 4.24 percent (0.000) for the advertising subscore. However, when we estimated regressions of the form in Table 7, but substituted the set of all three Q subscores for the overall Q-score in a single regression, the coefficient on the inventory subscore was not statistically

significant. This result suggests that much of the information about future returns in the inventory subscore is captured by the R&D and advertising subscores.

Control for Growth in Net Operating Assets

We also estimated regressions similar to those in Table 7 including both the Q-score and growth in net operating assets as explanatory variables to see whether the result might be due to growth in net operating assets (which affects the Q^A score). The coefficient on the growth variable was -0.066 (with a t-statistic of -2.42) and that on the Q-score was 0.218 (with a t-statistic of 2.70). Thus, Q continues to play an incremental role in forecasting stock returns, after controlling for growth in net operating assets.

Analysis of Q^A and Q^B

We also repeated the tests basing investment positions on either Q^A or Q^B . We were concerned that Q^B , a levels measure of the conservative accounting reserve (rather than a change variable like Q^A) might be correlated with risk factors or with other factors that predict returns. For example, reserves (and thus Q^B) depend on the accumulated amount of R&D investment relative to net operating assets and Chan et al. (2000) indicate that the market undervalues R&D. However, the zero-net-investment strategy using Q^A returned a size-adjusted return in Year +1 of 8.19 percent (significance level 0.000), and 1.41 percent (0.160) in Year +2, and the Q^B -based strategy earned 5.69 percent (0.000) in Year +1 and 2.37 percent (0.047) in Year +2.

VII. CONCLUSION

Commentators sometimes claim that the practice of conservatism in accounting produces higher-quality earnings. Conservatism yields lower earnings, they say, and so *prima facie* these conservative earnings are of higher quality. However, conservative accounting with investment growth depresses earnings and accounting rates of return, and creates unrecorded reserves. Firms slowing investment release these reserves, creating earnings and higher rates of return. If the change in investment is temporary, then the effects on earnings and rates of return are temporary, leading to lower-quality—or less sustainable—earnings.

Using a constructed index of accounting conservatism and a constructed index of the quality of earnings, this study identifies poor-quality earnings that result from changes in investment coupled with conservative accounting. The quality index forecasts how future core return on net operating assets is likely to change from current levels, so our index provides an analysis tool to identify low-quality earnings.

The study also shows that quality scores incrementally predict future stock returns, before transaction costs, over the 1976–1995 period, after controlling for measures commonly conjectured as risk proxies, and after controlling for growth in net operating assets and for accruals. These results suggest that the stock market did not penetrate the quality of earnings of firms with conservative accounting during the sample period. Accordingly, with the caveat that risk may not have been adequately controlled, the study indicates that a quality analysis of the form in the paper would have been rewarded with abnormal returns in the stock market during the sample period.

REFERENCES

- Baber, W., P. Fairfield, and J. Haggard. 1991. The effect of concern about reported income on discretionary spending decisions. *The Accounting Review* 66 (October): 818–829.
- Basu, S. 1997. The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics* 24: 3–37.

- Beaver, W., and E. Engel. 1996. Discretionary behavior with respect to allowances for loan losses and the behavior of security prices. *Journal of Accounting and Economics* 22: 177–206.
- , and S. Ryan. 2000. Biases and lags in book value and their effects on the ability of the book-to-market ratio to predict book return on equity. *Journal of Accounting Research* 38 (Spring): 127–148.
- Bernard, V., J. Thomas, and J. Wahlen. 1997. Accounting-based stock price anomalies: Separating market inefficiencies and risk. *Contemporary Accounting Research* (Summer): 89–136.
- Bublitz, B., and M. Ettredge. 1989. The information in discretionary outlays: Advertising, research, and development. *The Accounting Review* 64 (January): 108–124.
- Chan, L., J. Lakonishok, and T. Sougiannis. 2000. The stock market valuation of research and development expenditures. *Journal of Finance* 56 (December): 2431–2451.
- Fairfield, P., J. Whisenant, and T. Yohn. 2001. Accrued earnings and growth: Implications for earnings persistence and market mispricing. Working paper, Georgetown University.
- , and T. Yohn. 2001. Using asset turnover and profit margin to forecast changes in profitability. *Review of Accounting Studies* 6 (December): 371–385.
- Fama, E., and J. Macbeth. 1973. Risk, return and equilibrium—Empirical tests. *Journal of Political Economy* 81 (May/June): 607–636.
- , and K. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47: 427–465.
- Feltham, J., and J. Ohlson. 1995. Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research* 11 (2): 689–731.
- Freeman, R., and S. Tse. 1989. The multiperiod information content of accounting earnings: Confirmations and contradictions of previous earnings reports. *Journal of Accounting Research* 27 (Supplement): 49–79.
- Gjesdal, F. 1999. A steady state growth valuation model: A note on accounting and valuation. Working paper, Norwegian School of Economics, Bergen, Norway.
- Greenball, M. 1969. Appraising alternative methods of accounting for accelerated tax depreciation: A relative accuracy approach. *Journal of Accounting Research* 7: 262–289.
- Hall, B. 1993. Industrial research during the 1980s: Did the rate of return fall? *Brookings Papers on Economic Activity, Microeconomics*: 289–344.
- Healy, P. 1985. The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics* 7: 85–107.
- , and J. Wahlen. 1999. A review of the earnings management literature and its implications for standard setting. *Accounting Horizons* 13 (December): 365–383.
- Jones, J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29 (Autumn): 193–288.
- Lee, C., J. Myers, and B. Swaminathan. 1999. What is the intrinsic value of the Dow? *Journal of Finance* 54: 1693–1741.
- Lev, B., and T. Sougiannis. 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics* 21: 107–138.
- Moehrl, S. 2002. Do firms use restructuring charge reversals to meet earnings targets? *The Accounting Review* 77 (April): 397–414.
- Mulford, C., and E. Comiskey. 1996. *Financial Warnings*. New York, NY: John Wiley & Sons.
- Nissim, D., and S. Penman. 2001. Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies* 6: 109–154.
- Ou, J., and S. Penman. 1989. Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics* 14: 295–329.
- Penman, S., and T. Sougiannis. 1998. A comparison of dividend, cash flow, and earnings approaches to equity valuation. *Contemporary Accounting Research* 15: 343–383.
- . 2001. *Financial Statement Analysis and Security Valuation*. New York, NY: McGraw-Hill.
- Perry, S., and R. Grinaker. 1994. Earnings expectations and discretionary research and development spending. *Accounting Horizons* (December): 43–51.

- Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71: 289–315.
- Teoh, S., T. Wong, and G. Rao. 1998. Are accruals during initial public offerings opportunistic? *Review of Accounting Studies* 3: 175–208.
- Zhang, X. 2000. Conservative accounting and equity valuation. *Journal of Accounting and Economics* 29: 125–149.
- . 2001. Accounting conservatism and the analysis of line items in earnings forecasting and valuation. Working Paper, University of California, Berkeley.

An Experimental Investigation of Self-Serving Biases in an Auditing Trust Game: The Effect of Group Affiliation

Ronald R. King
Washington University

ABSTRACT: I report the results of an experiment designed to investigate the influence of noncredible communications and group affiliation on auditors' formation of self-serving bias. I find that manager-subjects use noncredible communications to induce auditors to develop an unwarranted trust of managers (i.e., a biased judgment). However, the bias is neutralized when auditor-subjects belong to groups that create social pressure to conform to group norms. Thus, my finding calls into question the Bazerman et al. (1997) conclusion that auditors cannot conduct impartial audits due to self-serving biases resulting from repeated interactions between auditors and their clients.

Keywords: *self-serving bias; auditor independence; trust; groups.*

Data Availability: *Data are available from the author.*

[W]e maintain that audit failures are the natural product of the auditor-client relationship. Under current institutional arrangements, it is psychologically impossible for auditors to maintain their objectivity; cases of audit failure are inevitable, even with the most honest auditors.

—Bazerman et al. (1997)

Helpful comments were received from Lynnea Brumbaugh-Walter, Bryan Church, John Dickhaut, Nick Dopuch, Mahendra Gupta, Steven Kachelmeier, Kathryn Kadous, Brian Mayhew, Claus Langfred, Bill Messier, Mark Peecher, Bill Rankin, P. K. Sen, Shyam Sunder, Ping Zhou, two anonymous reviewers, seminar participants at Georgia State University, The Ohio State University, University of Illinois Auditing Symposium, Washington University, 2001 Midyear Auditing meetings, and the First Asian Conference on Experimental Business Research held at Hong Kong University of Science and Technology. Research assistance was received from Amy Choy and Jim Holloway. Financial support from the Taylor Experimental Laboratory at the Olin School of Business is gratefully acknowledged.

Submitted July 2000
Accepted September 2001

I. INTRODUCTION

In a prominently cited commentary (Securities and Exchange Commission 2000), Bazerman et al. (1997) argue that auditors suffer from an unconscious *self-serving bias*, and thus cannot conduct impartial audits. The bias arises because of auditors' repeated close interactions with client management and their limited and distant interactions with investors.¹ Unconscious bias is particularly problematic because economic sanctions are less likely to mitigate it. In this paper, I report the results of an experiment designed to investigate the extent to which group affiliation, a noneconomic factor, can reduce self-serving biases in an audit trust game.

Although provocative, the Bazerman et al. (1997) conclusion is premature for two reasons: first, the experiments on which they base their conclusion are not oriented to auditing settings;² and second, the authors do not consider whether auditors' psychological bonds created by group membership may reduce self-serving biases. Auditors generally belong to engagement teams, audit firms, and professional organizations, all of which can create psychological affiliations that may mitigate biases that auditors may form due to repeated close interactions with client personnel.

I investigate self-serving biases in an experimental setting that has auditing frictions, while also incorporating the precepts of experimental economics. I investigate an "audit trust" game played by a pair of subjects: a manager and an auditor. Each manager chooses a level of misappropriation, hereafter denoted as *fraud*. The auditor's goal is to predict the manager's fraud level, because correctly anticipating the fraud level leads to an effective audit.

I develop a 2×2 full factorial, between-subject experimental design. The first treatment manipulates the manager's communications to the auditor (referred to as the *puffery treatment*). The puffery treatment has two levels—the no-puffery setting and the yes-puffery setting. In the no-puffery setting, managers cannot communicate to their auditor-partners their nonbinding intention to commit low levels of fraud. Without such communication, auditors have a limited basis on which to trust managers, and economic theory predicts that auditors will protect themselves by selecting the *Nash strategy*, the best response to the expected strategy of the other player. In the yes-puffery settings, managers may communicate their intentions to commit low levels of fraud by using nonbinding communication mechanisms, or *puffery*. That is, a manager may "promise" his auditor-counterpart that he intends to cooperate by committing low levels of fraud, but not implement the promise. The puffery treatment captures the psychological links developed between auditors and their clients with repeated interactions that do not necessarily reflect the client's true trustworthiness.

Managers' use of puffery may lead auditors to form biased judgments about managers' trustworthiness. Economic theory predicts that auditors should ignore puffery because they can anticipate managers' strategic opportunities. However, research on "cheap talk" (e.g., Cooper et al. 1989) suggests that puffery can create familiarity between players and thus facilitate the formation of reciprocal trust. I enhance managers' appreciation of the strategic

¹ Several characteristics of auditing services increase auditors' propensity to develop biases. Managers can apply immediate economic and noneconomic pressures against auditors who disagree with them, whereas economic penalties for audit errors are uncertain and delayed long into the future (Palmrose 1994). In addition, ambiguous financial-reporting standards allow auditors to rationalize a range of decisions (Libby et al. 2001).

² Bazerman et al. (1997) draw their conclusions about the impossibility of auditor independence based on a number of research articles that found subjects who were assigned different roles, such as plaintiff or defendant, made different decisions on how a judge would decide a case. See Babcock and Loewenstein (1997) for a review of this research.

opportunities available to them by (1) providing managers a separate instructional session in which they play the roles of both auditor and manager, thus learning the vulnerabilities of auditors who trust managers, (2) having managers participate in multiple sessions, and (3) explaining to managers how they might use puffery to create trust and then act opportunistically. The potential incentives for opportunism in my experimental setting are not intended to correspond to levels that exist in naturally occurring settings. Rather, I design the settings to create the potential for auditor bias in order to investigate the extent to which group affiliation can undo it.

The second treatment manipulates the strength of group interactions among auditors (referred to as the *group treatment*). The group treatment has two levels—the weak group setting and the strong group setting. In the weak group setting, interactions occur only between auditors and managers, not between different auditors. In the strong group setting, auditors become members of a group. Group membership can create pressures to conform to group norms, which in turn increases the psychological cost of penalties. In addition, group membership can create a team identity that motivates auditors to focus more on the collective goal of conducting appropriate audits. Although psychological costs increase in the strong group setting, economic incentives do not change. These two treatments allow me to investigate (1) whether managers' noncredible communications exacerbate auditors' susceptibility to bias (as Bazerman et al. [1997] imply); and (2) the extent to which group affiliation can neutralize auditors' bias. This design is consistent with Kachelmeier's (1996a, 1996b) recommendation that experimental research incorporate both economic and psychological forces.

I report the decisions of 44 manager/auditor pairs: 11 pairs in each of the four settings (no-puffery/weak group; yes-puffery/weak group; no-puffery/strong group; yes-puffery/strong group). University students participated as subjects. The results indicate that managers' use of puffery in the yes-puffery settings significantly increases auditors' willingness to trust managers, relative to the no-puffery settings. Auditors in the strong group settings trust managers significantly less than did auditors in the weak group settings. I also find that the self-serving bias does not differ significantly between the yes-puffery/strong group and the no-puffery/weak group settings. Thus, psychological forces associated with group affiliation can neutralize the bias, calling into question Bazerman et al.'s (1997) conclusion that auditors cannot conduct impartial audits due close repeated interactions with their clients.

In the next section I discuss the game that the subjects played. Section III discusses the experimental method, and the development of hypotheses follows in Section IV. The results are presented in Section V, and Section VI provides a discussion of the research.

II. GAME

There are two players: a manager and an auditor. The manager selects a fraud level, which determines the amount of resources the manager attempts to embezzle.³ For any given fraud level, the manager faces one of two outcomes. The first outcome occurs if the auditor does not detect the fraud, so the manager's payoff is relatively high because he retains the embezzled resources. In the second outcome, the auditor detects the fraud, so the manager's payoff is relatively low because he is sanctioned. The auditor's objective is to select an audit level to match the manager's fraud level because a correct match leads

³ I present the game in a fraud context; however, characterizing the game as a reporting decision would allow the same frictions. See Bloomfield (1995) and Matsumura and Tucker (1992) for a discussion of modeling issues in fraud settings.

to an effective audit. For any given audit level, the auditor faces one of two outcomes. In the first, the auditor's payoff is relatively high because the auditor is not assessed a penalty. In the second outcome, the auditor's payoff is relatively low because he is penalized for not detecting the fraud.

Each player has three choice options in the experiment. This simplifies the game and provides clearer demarcations for comparisons across the settings. Table 1 shows the payoff matrix. The manager's choices appear in the columns, with the three fraud levels denoted as *cooperate* (lowest level of fraud), *Nash* (middle level of fraud), and *cheat* (highest level of fraud). The auditor's choices appear in the rows, denoted as *trust* (lowest audit; with the

TABLE 1
Payoff Matrix Used in the Experiment^a

		<i>Manager's Fraud Level</i>		
		<i>Manager Selects the Cooperate Fraud Level</i>	<i>Manager Selects the Nash Fraud Level</i>	<i>Manager Selects the Cheat Fraud Level</i>
Auditor Selects the Trust Audit Level	Manager's payoffs			
	Expected payoff	46.82	58.76	65.96
	Fraud detected (payoff; probability)	-30.43; 0.29	-68.64; 0.29	-113.75; 0.29
	Fraud not detected (payoff; probability)	78.34; 0.71	110.74; 0.71	139.29; 0.71
	Auditor's payoffs			
	Expected payoff	22.32	8.51	-3.67
	No penalty (payoff; probability)	53.61; 0.81	53.61; 0.81	53.61; 0.81
	Penalty (payoff; probability)	-111.42; 0.19	-184.29; 0.19	-248.49; 0.19
	Manager's payoffs			
	Expected payoff	32.89	35.78	33.55
Auditor Selects the Nash Audit Level	Fraud detected (payoff; probability)	-30.43; 0.42	-68.64; 0.42	-113.75; 0.42
	Fraud not detected (payoff; probability)	78.34; 0.58	110.74; 0.58	139.29; 0.58
	Auditor's payoffs			
	Expected payoff	21.08	9.75	-0.22
	No penalty (payoff; probability)	46.73; 0.84	46.73; 0.84	46.73; 0.84
	Penalty (payoff; probability)	-118.31; 0.16	-191.17; 0.16	-255.37; 0.16
	Manager's payoffs			
	Expected payoff	20.62	15.55	5.01
	Fraud detected (payoff; probability)	-30.43; 0.53	-68.64; 0.53	-113.75; 0.53
	Fraud not detected (payoff; probability)	78.34; 0.47	110.74; 0.47	139.29; 0.47
Auditor Selects the Defensive Audit Level	Auditor's payoffs			
	Expected payoff	17.92	8.79	0.74
	No penalty (payoff; probability)	38.60; 0.87	38.60; 0.87	38.60; 0.87
	Penalty (payoff; probability)	-126.44; 0.13	-199.30; 0.13	-263.50; 0.13

^a The nine audit/fraud combinations show the expected payoffs for both the manager and auditor, and the payoff and probability components that produce the expected payoffs.

lowest cost, and the lowest probability of detecting fraud), *Nash* (middle audit), and *defensive* (highest audit). The three options for each player lead to nine possible audit/fraud combinations. The top part of each cell shows the manager's expected payoff along with the two possible payoffs and the associated probability of each payoff. The bottom part of each cell shows the auditor's payoffs and probabilities. I constructed the payoff matrix by specifying a model that incorporates the frictions discussed below. The details of the model are available upon request.

Suppose an auditor selects the trust audit and the manager selects the cooperate fraud level (trust/cooperate combination). Under this combination, the manager's expected payoff is 46.82, which is a weighted average of two possible payoffs. Specifically, when the auditor detects fraud, the manager receives a payoff of -30.43 (i.e., he pays 30.43). This occurs 29 percent of the time under this choice combination. The manager receives a payoff of 78.34 when the auditor does not detect fraud, which occurs 71 percent of the time under this combination. Similarly, under the trust/cooperate combination, the auditor's expected payoff of 22.32 is based on two outcomes. In the first, the auditor is not penalized, which occurs 81 percent of the time and the auditor receives 53.61. When the auditor is penalized, he receives a payoff of -111.42 (i.e., he pays 111.42), which occurs 19 percent of the time under this combination. The experimental parameters, such as the probabilities of fraud and the probabilities of detecting fraud, do not necessarily represent real-world levels. Rather, these parameters achieve separation in payoffs across the choice options, thereby enhancing the experiment's internal validity.

Characteristics of Manager Payoffs

The Manager's Possible Payoffs Depend on His Selected Fraud Level, but Not on the Audit Level

Consider the manager's possible payoffs when he selects the cooperate fraud level (column 1). Under this fraud level, the manager's payoffs are either -30.43 or 78.34. These payoffs are constant across the three audit levels, although the expected payoffs change because higher audit levels increase the probabilities that the auditor detects the fraud. The two possible payoffs are independent of the audit level because the audit fee is fixed—simplifying the setting.

When the manager commits higher levels of fraud, his range of possible payoffs expands. Under the Nash fraud level, the manager's payoffs are -68.64 or 110.74, and under the cheat fraud level they are -113.75 or 139.29. These differential payoffs across fraud levels reflect that low levels of fraud result in smaller sanctions against the manager when the auditor *detects* fraud and also that low levels of fraud provide the manager a relatively small payoff if the auditor does *not detect* the fraud (lower risk; lower reward).

The Audit Determines the Probability of Each Payoff for the Manager, but the Probabilities Are Constant across Fraud Levels

Under the trust audit level, the manager faces sanctions 29 percent of the time across all three fraud levels. Constant percentages across fraud levels imply that audit effectiveness does not depend on the size of the fraud; that is, more fraud does not increase the probability of detection. However, the probability auditors detect fraud increases with the audit level. This, in turn, raises the probability of managerial sanctions. The auditor detects the manager's fraud 29 percent of the time under the trust audit, increasing to 42 percent and 53 percent for the Nash and defensive audits, respectively.

Characteristics of Auditor Payoffs

The Audit Level Determines the Probability That the Auditor Pays a Penalty

The probability that the auditor pays a penalty for failing to detect fraud decreases with higher audit levels. Under the trust audit, the auditor is penalized 19 percent of the time. The percentage decreases to 16 percent and 13 percent for the Nash and defensive audits, respectively. Again, these percentages may not correspond to real-world probabilities of auditor sanctions; they achieve a meaningful payoff separation across the choice options.

The Manager's Fraud Level Determines the Auditor's Payoffs Only When the Auditor Is Penalized

To see this, consider again the trust/cooperate combination. The auditor's payoffs under this audit level are either -111.42 (auditor is penalized) or 53.61 (auditor is not penalized). If the auditor is not penalized, then the payoffs do not change when the manager commits higher levels of fraud. Thus, the 53.61 payoff remains the same for the cooperate, Nash, and cheat fraud levels. However, when the auditor is penalized, the payoffs change to -184.29 and -248.49 for the Nash and cheat fraud levels, respectively. The auditor's penalty increases with higher fraud levels because it is a function of the amount of undetected fraud. Note that the auditor's payoffs decrease with higher audit levels regardless of whether the auditor is penalized, because higher audit levels are more costly.

Summary of Payoff Matrix

The manager's expected payoffs depend on his fraud level (which determines the payoff amounts) and the audit level (which determines the probabilities of each payoff). The auditor's expected payoffs depend on the fraud level (when the auditor is penalized) and his audit level (which affects the cost of producing the audit and the probabilities of each payoff). An inspection of the payoff matrix shows that the manager has incentives to cheat when expecting the auditor to trust; to select the Nash audit when expecting the auditor to select the Nash audit; and to cooperate when expecting the defensive audit. Auditors prefer the trust audit when expecting the manager to cooperate; the Nash audit when expecting the Nash fraud level; and the defensive audit when expecting the manager to cheat. Thus, the manager's best response is the off-diagonal and the auditor's best response is the diagonal.

The Nash equilibrium of this audit trust game is the center cell of the matrix (Nash audit/Nash fraud). This outcome represents the solution in which each player best responds to his or her anticipation of the other player's choice, and neither party has a unilateral incentive to change his or her choice. The expected payoffs to the auditor and to the manager are 9.75 and 35.78 , respectively. Although this Nash equilibrium implies risk neutrality, the behavioral hypotheses discussed below are not sensitive to this assumption.⁴

Note that in the trust/cooperate combination, the payoffs to the auditor and manager are 22.32 and 46.82 , respectively. This outcome Pareto dominates the Nash equilibrium, meaning that both parties are better off under this combination than under the Nash combination. This possibility creates a fertile environment for self-serving bias because auditors may hold biased perceptions of the manager's truthfulness, hoping to achieve Pareto superior payoffs. However, under the trust/cooperate combination, the manager has incentives

⁴ Risk-averse managers would be predicted to decrease fraud levels, and risk-averse auditors would increase the audit levels, *ceteris paribus*. See Kachelmeier (1991) for a discussion of risk neutrality; see Bloomfield (1997) for an example of research that investigates settings that use expected values in the game; see King and Schwartz (2000) for an example that has risky payoffs; and see Mayhew et al. (2000) for a study that induces risk levels.

to cheat and thus increase his expected payoff to 65.96, which reduces the auditor's expected payoff to -3.67 . The auditor has no incentive to change from the trust/cooperate combination. Thus, under this asymmetric payoff structure, auditors benefit from having trustworthy clients. However, the manager has unilateral incentives to violate the auditor's trust, which creates risks for the auditor.

III. EXPERIMENTAL METHOD

Subjects are business-school student volunteers, paid based on the outcomes of their choices. Either 10 or 12 subjects participate in each session, half playing the role of the manager and half playing the role of the auditor. I conduct two sessions under each of the four settings, which include the no-puffery/weak group setting (denoted NP/WG); the yes-puffery/weak group setting (YP/WG); the no-puffery/strong group setting (NP/SG); and the yes-puffery/strong group setting (YP/SG). These eight sessions produce 11 subject-pairs for each of the four settings, for a total of 44 pairs.

I conduct each experimental session on networked personal computers with privacy maintained, except as noted below. The instructions, which are available upon request, do not use terms with real-world connotations, such as *fraud* or *auditor*. Rather, managers are referred to as "A" players, and auditors are referred to as "B" players. Haynes and Kachelmeier (1998) argue that context can affect the behavior of subjects in ways that are extraneous to the study's theoretical predictions.

All managers have special training several days prior to the actual experimental sessions, which helps them appreciate their strategic opportunities. During the training session, managers play both roles in a series of practice rounds and were informed of the strategic possibilities inherent in the game. The trained managers participated in multiple sessions of the experiment.

In the experiment, auditors arrive at the experimental laboratory one-half hour earlier than managers to receive their instructions. When the managers arrive, the instructions are summarized for both managers and auditors to enhance common knowledge of the game. Auditors are aware that managers had previous experience in the experiment. The purpose of having experienced managers is to create an environment in which managers have information and experience advantages and to increase managers' awareness of the potential for opportunistic behavior. I did this to create an environment ripe with the potential for auditors to fall prey to self-serving biases in order to investigate the extent to which groups can mitigate the bias.⁵

Table 2 summarizes how the treatments were operationalized in the experiment. The first column identifies the four experimental settings. The second column indicates how I operationalize the WG and SG settings. In the WG settings, auditors receive instructions collectively, but without any indication that they are part of a group. In the SG settings, I establish group cohesion/identity by asking auditors to introduce themselves to the other auditors in the session, to use nametags, to select a team name, and to work together on the quiz.⁶ Auditors are also told that the name of the group member with the highest penalty level at the end of the session will be posted on the board for all to view. This treatment is entirely psychological in that there are no explicit economic ties among the auditor group members.

⁵ I randomly assigned managers to settings and investigated the possibility of order or experience patterns on the part of managers. I found no evidence to indicate any such patterns.

⁶ Auditors could discuss the quiz with each other in the SG setting, so the SG settings may have increased auditors' understanding of managers' strategic options. However, auditors' average scores on quizzes, which measured their understanding of the game, did not differ between the WG and SG settings.

TABLE 2
Summary of the Approach Used to Operationalize the Treatments

<i>Experimental Setting</i>	<i>Procedures for Auditor Group Treatment^a</i>	<i>Phase 1 Puffery-Action Phase: 5 Rounds^b</i>	<i>Phase 2 Regular Play Phase: 18 or 19 Rounds</i>	<i>Phase 3 Manager's Actions Disclosed to the Auditor at the End of Regular Play Phase and Resolution Phase: 6 Rounds</i>
NP/WG (no-puffery/weak group)	No	No	No puffery	Yes
YP/WG (yes-puffery/weak group)	No	Yes	Puffery	Yes
NP/SG (no-puffery/strong group)	Yes	No	No puffery	Yes
YP/SG (yes-puffery/strong group)	Yes	Yes	Puffery	Yes

^a "No" refers to settings where interactions occur only between auditors and managers, not between different auditors. "Yes" refers to settings where auditors are formed into groups.

^b "No" refers to settings where managers could not communicate their intentions to cooperate using nonbinding communication mechanisms. "Yes" refers to settings where managers could communicate their intentions to cooperate using nonbinding communication mechanisms; auditors observed the managers' fraud choices after this stage of the experiment.

The third and fourth columns indicate how settings differ with respect to the puffery treatment. Only subjects in the YP settings complete Phase 1: the puffery-action phase, where each manager and auditor pair interacts for five rounds and each party receives a fixed payment of \$0.50 per round. Subjects are informed that the fixed payment is independent of the decisions they make during this phase. At the end of the fifth round, each auditor observes the corresponding manager's fraud choices made during this phase. Thus, managers can form reputations for cooperation by selecting the cooperate option (hence the "puffery-action" designation). In addition, during Phase 1, managers can use puffery-talk (discussed below). The purpose of the puffery-action phase is to enhance trust between managers and auditors. Managers' instructions encouraged them to form reputations for honesty during the puffery-action phase by informing them of the strategic possibilities associated with having the auditors view their willingness to cooperate during this phase. To further facilitate this trust, the puffery-action phase is referred to as the "Get Acquainted" session for the subjects. This treatment surrogates for interactions between clients and auditors that do not necessarily reflect the client's true trustworthiness. Importantly, auditors understand that managers' puffery is not binding; that is, a manager may promise to cooperate but actually select the Nash or cheat fraud choice.

In Phase 2: regular play, subjects earn money based on the outcomes of their joint decisions. This phase consists of 18 or 19 rounds of play. In the YP settings, managers can engage in puffery during the regular play phase—that is, they may send their auditor-partners the message, "I will choose option 1 this round" (discussed in more detail below).

If they elect not to send this message, then the message “No message this round” is sent. This restriction limits managers’ communication options so that they can communicate only their willingness to cooperate (rather than creating communication noise that may be difficult to decipher *ex post*).

An auditor learns when his corresponding manager is sanctioned at the end of each round, but not the fraud level, which the auditor does not learn until the end of the regular play phase. This delays when auditors learn of the managers’ trustworthiness, consistent with delayed feedback in naturally occurring settings (Palmrose 1994). The auditor knows that he will not pay a penalty for rounds when a manager is sanctioned; however, this sanction information tells the auditor nothing about the manager’s fraud level because the frequency of sanctions depends entirely on the audit level, a feature known by the auditor (see Table 1).

In contrast, managers who are sanctioned *do* learn the audit levels. However, auditors do not know that the managers learn this. I chose not to inform auditors that their managers receive information about audit levels (when the auditor detects fraud) to lessen auditors’ ability to signal the audit level to managers and thus become a leader, as in a leader/follower game.⁷ Thus, auditors’ information restrictions limit them from forming a justifiable basis for trusting managers. These conditions create a fertile environment for self-serving biases to arise.

At the end of the regular play phase, each auditor learns of his corresponding manager’s choices, but not of the actual penalties. This procedure captures the extent to which auditors modify their behavior after they observe the strategies of their manager partners, but without knowing the consequences of the audit and fraud choices. Each pair then continues for another six rounds of play, denoted as the resolution phase (shown as Phase 3 in Table 2).⁸ At the end of the resolution phase, auditors learn their penalties, their net earnings for each round, and their total cash earnings (earnings average approximately \$25 for a two-hour time commitment). For SG settings, the administrator posts the name of the auditor with the largest cumulative penalties. Finally, the administrator pays subjects privately and dismisses them.

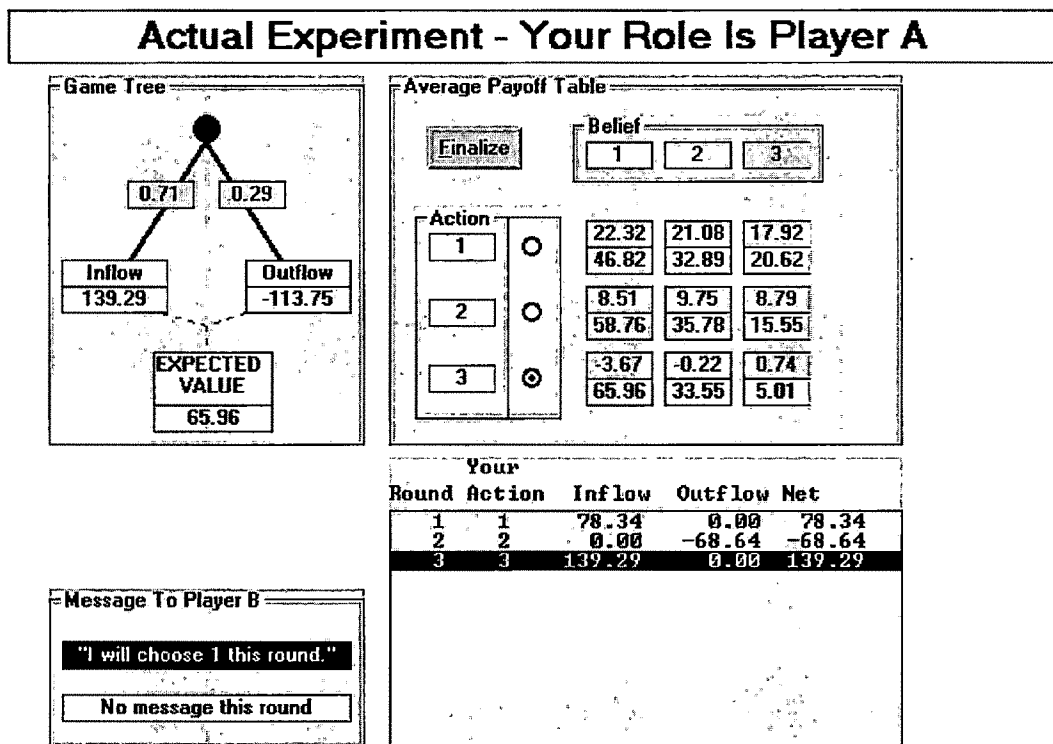
Figure 1 displays the YP/SG setting screens for managers. I denote the manager as Player A and the auditor as Player B. The screen for managers has four windows: the “Game Tree” window (upper left-hand side), the “Average Payoff Table” window (upper right-hand side), the “Message to Player B” window (bottom left-hand side), and the “history” window (bottom right-hand side). Subjects use the average payoff table window to select their choices. This window also shows the expected payoff matrix for all nine audit/fraud combinations (the experimental materials refer to audits as beliefs). In Figure 1, the manager has selected Action 3 (cheat) and Belief about the audit level 1 (trust). The Average Payoff Table portion of the screen highlights the expected payoffs for both players, with the payoffs to the auditor (manager) in the top (bottom) row so each player can see the other player’s expected payoffs for each combination.

The Game Tree window in the upper left-hand side of the screen shows the probabilities and the amounts for the audit/fraud combination selected in the Average Payoff Table. As noted above, the manager has chosen the trust/cheat combination. The left branch of the

⁷ This approach does not represent deceit, merely that the information was not commonly known (Roth 1987). An interesting question is whether auditors can be leaders in a leader/follower game in practice; however this issue is beyond the scope of this study.

⁸ The resolution phase has only six rounds (and the regular play has 18 or 19), so that managers are not motivated to use the regular play phase to form reputations to be consumed in the resolution phase. Auditors do not know the number of rounds, but do know there are fewer rounds in the resolution phase than in the regular play phase.

FIGURE 1
Manager's Computer Screen under Yes-Puffery/Strong Group Setting



Game Tree shows the payoffs to the manager if the auditor does not detect fraud (denoted as *Inflow*). The probability of no sanction is 0.71, and the payoff to the manager is 139.29. The right branch shows that the probability of fraud detection (denoted as *Outflow*) is 0.29 and that the payoff to the manager is -113.75. The expected value of these two possibilities is 65.96. The screen updates the Game Tree each time a subject chooses a different audit/fraud combination.

The Message to Player B window on the bottom left appears only in YP settings. This window allows the manager to exercise his option to communicate a nonbinding intent to cooperate (Action 1). Auditors know that managers' communications are nonbinding. The History window shows the history of play for past rounds. In Figure 1, the manager selected Action 1 in round 1, which generated a payoff of 78.34 (the manager was not sanctioned). In round 2, the manager chose Action 2 and was sanctioned, paying 68.64 for that round. In round 3, the manager chose Action 3 and was not sanctioned, receiving 139.29. Auditors' screens had the same format as managers' screens; a copy is available upon request.

IV. HYPOTHESES

The first set of hypotheses deals with auditors' choices in the NP/WG setting to establish a benchmark. The second set of hypotheses compares the NP settings to the YP settings and the third set compares the WG settings with the SG settings. The last set of hypotheses compares regular play rounds to resolution rounds.

H1: Hypotheses for the NP/WG (Baseline) Setting

This set of hypotheses addresses audit levels during the regular play phase of the NP/WG setting. I set up two alternatives, one based on the Nash prediction and the second on the trust/reciprocity literature. Under the Nash equilibrium, auditors analyze the expected payoffs for both parties and recognize managers' incentives to defect from a choice to cooperate. The limited information available to auditors should constrain them from forming a justifiable basis for trusting managers. Without a basis for trust, economic theory predicts that auditors will protect themselves by selecting the Nash audit. Theoretically, they should choose the Nash level 100 percent of the time, but previous research shows that such an extreme prediction is unsubstantiated. For example, Bloomfield (1995, 1997) finds that equilibria have low predictive power when each player's optimal strategy is highly sensitive to the other's choice, as in my setting. Thus, I test whether auditors choose the Nash level more frequently than the other audit levels.

The second alternative, based on the trust/reciprocity literature, predicts that auditors select a relatively higher percentage of trust audits. This outcome is optimal for the auditor only if the manager cooperates. The rationale for the trust alternative is based on previous economic experiments that indicate subjects trust others to reciprocate more than predicted by economic theory in noncooperative games (Berg et al. 1995; Dawes and Thaler 1988). As stated above, the trust/cooperate pair Pareto dominates the Nash equilibrium. The two competing hypotheses for NP/WG setting are stated below.

H1 (Nash): Auditors select the Nash audit more frequently than the other audit levels in the NP/WG setting.

H1 (Trust): Auditors select the trust audit more frequently than the other audit levels in the NP/WG setting.

H2: NP Settings Compared to YP Settings

I set up three alternatives, the first based on the Nash prediction, and the second and third based on the "cheap talk" and self-serving bias literatures. The Nash model predicts that puffery should not affect audit levels differentially in the NP and YP settings because auditors understand managers' incentives to use puffery opportunistically. Thus, auditors are predicted to select the Nash audit equally in the NP and YP settings.

The cheap talk literature has shown that nonbinding communication can enhance mutual cooperation in coordination games where both parties benefit from cooperation (Cooper et al. 1989). In addition, research has shown that certain types of cheap talk can enhance cooperation in a Prisoners' Dilemma game where players have symmetric payoffs and both can communicate an intent to punish a defector. Based on these findings, puffery in my setting may induce auditors to perceive their managers as trustworthy. This, in turn, will induce auditors to select the trust audit. The self-serving bias (SSB) prediction is that auditors' trust is not reciprocated by managers, who use puffery opportunistically. The Pareto Optimal (PO) alternative predicts that auditors are trusting (as discussed in the SSB alternative) and that managers reciprocate the trust, raising the payoffs for both players. The null and the alternatives are stated below.

H2 (Nash): The frequency of Nash audits will be the same in the NP and YP settings.

H2 (SSB): Auditors select the trust audit more frequently in the YP settings than in the NP settings *and* managers do not reciprocate by lowering the fraud level to match the lower audit level (thus leading to a self-serving bias on auditors' part).

H2 (PO): Auditors select the trust audit more frequently in the YP settings than in the NP settings *and* managers reciprocate by lowering the fraud level to match the lower audit level (thus leading to a Pareto improvement over Nash).

H3: WG Settings Compared to SG Settings

The economic incentives are the same across the SG and WG settings. Thus, if auditors are not affected by the psychological influences of the SG treatment, then their choices should not differ across the SG and WG settings, as predicted by H3 (Null) shown below.

The alternative hypothesis predicts that the SG treatment creates psychological forces that reduce auditors' trust of managers in the SG settings relative to the WG settings. There are two possible forces inherent in the SG treatment. First, auditors may attach a psychological cost from seeing their name posted publicly and, thus, the treatment may induce auditors to increase audit levels in an attempt to avoid this psychologically costly outcome. That is, auditors increase their audit level to avoid a potential "loss of face" associated with the public identification created by mistakenly trusting their manager-partner (Kachelmeier and Shehata 1997). The second reason the SG treatment may reduce auditors' trust of managers is that the SG treatment may create a team identity that motivates auditors to focus more on their collective goal of conducting appropriate audits than on their individual self-interested motives (Abrams and Hogg 1999). Both of these psychological forces may reduce auditors' trust of managers in the SG setting (although I cannot disentangle the relative influences of the two forces). The null and competing alternatives are stated below.

H3 (Null): Audit levels will be the same in the WG and SG settings.

H3 (Group Effect): Auditors select the Nash or the defensive audit more frequently in the SG settings than in the WG settings.

H4: Comparing Audit Levels across the Resolution and Regular Play Phases

Each auditor learns of his corresponding manager's fraud levels at the end of the regular play phase, but does not learn of his own consequent penalties. Hypothesis 4 investigates the effect of this information on auditor choices during the resolution phase (see Table 2). The null hypothesis predicts no difference between audit levels across the regular play rounds and those from the resolution rounds, suggesting that the information about the manager's fraud level does not change the auditor's strategy. That is, audits are in "steady state" during the regular play phase (see Bloomfield [1997] for a discussion of equilibrium dynamics). The alternative predicts that an auditor will be less likely to choose the trust audit after learning the manager's fraud level. The alternative is based on the fact that auditors have no information about the managers during the regular play rounds, and thus may be overly trusting of managers during the regular play phase. After learning of the manager's fraud choices, auditors become less trusting.

H4 (Null): Audit levels will be the same across the regular play and the resolution phases.

H4 (Alternative): Audit levels will indicate less trust during the resolution phase than in the regular play phase.

V. RESULTS

My analysis focuses on choices made by auditor/manager pairs. There are 11 auditor-manager pairs in each setting, generating a total of 44 pairs.⁹ I create three measures to investigate the hypotheses: the audit index, the fraud index, and payoff difference index. I calculate the audit index by assigning the value of 0 to the trust audit, 1 to the Nash audit, and 2 to the defensive audit. I summed these values over the total number of periods (18 or 19) and then divided by the number of periods. This value is then multiplied by 100 to yield the audit index for each auditor. For example, consider an auditor who selected the trust audit nine times, the Nash audit six times, and the defensive audit three times, out of 18 total choices. The audit index of 66.67 is calculated as $((9 \times 0 + 6 \times 1 + 3 \times 2)/18)$. A lower index score indicates the auditor is more trusting of the manager, and a high index indicates the auditor is less trusting.

In like fashion, I construct the manager's fraud index by assigning a value of 0 to the cooperate fraud level, 1 to the Nash fraud level, and 2 to the cheat fraud level. Thus, a lower index score indicates the manager is more cooperative, and a high score indicates a greater propensity to cheat.

The third measure is the *ex post* payoff difference index. This measure applies to each auditor/manager pair and is calculated by subtracting the auditor's total *ex post* expected payoffs from the manager's expected payoff (based on audit and fraud-level combinations for each party, as in Table 1).

Table 3 summarizes the three measures across the four settings for the regular play phase. Column 2 shows data for the NP settings and column 3 shows data for the YP settings. The differences between the two puffery conditions appear in column 4. For example, the number 29.77 in column 4 for the NP/WG setting represents the difference between the average audit index across the NP/WG and YP/WG settings ($82.14 - 52.37$). The weak group and strong group settings appear in the rows of the table, with the bottom three rows showing the differences in the measures between the weak and strong group treatments.

Results for H1: Hypotheses for the NP/WG Setting (Baseline)

This hypothesis investigates audit levels in the NP/WG setting. The two alternatives are the Nash and the trust predictions. Panel A of Figure 2 displays the aggregate data for the combined audit/fraud choices in the NP/WG setting. The figure indicates the trust audit is the auditors' modal choice (selected 42 percent of the time) and the cheat option is the modal choice for managers (selected 46 percent of the time). The results indicate a mismatch that is detrimental to the auditors. (Hypothesis 2, discussed shortly, provides a more rigorous test of self-serving bias by examining the incremental influence of nonbinding communications.)

Table 3 shows that the average audit index was 82.14, a level indicating a tendency toward the trust audit choice. Using approximate randomization, I find no evidence that auditors' choices are predominately Nash ($p > 0.10$) or predominately trust ($p > 0.10$).¹⁰

⁹ In the NP/WG, YP/WG, and YP/SG settings, five pairs of subjects interacted for 19 rounds, and six pairs interacted for 18 rounds during the regular play phase. This generated a total of 203 choices $[(5 \text{ pairs} \times 19 \text{ rounds}) + (6 \text{ pairs} \times 18 \text{ rounds}) = 203]$ for these three settings. In the NP/SG setting, all 11 pairs had 18 rounds during the regular play phase, thus producing 198 choices $(11 \text{ pairs} \times 18 \text{ rounds} = 198)$. An Appendix is available upon request that details individual subjects' choices.

¹⁰ The approximate randomization approach generates a test statistic from the actual data, then the data are combined and repeatedly partitioned and the statistic is recomputed for each of the random partitions. The p-values represent the percentage of times the statistic from 40,000 random partitions is less than the test statistic for the actual data. Edgington (1987) discusses the approximate randomization method.

TABLE 3
Average Audit Index, Average Fraud Index, and Average Payoff Difference for Regular Play Phase^a

	<i>No-Puffery Setting</i>	<i>Yes-Puffery Setting</i>	<i>Differences of Measures across the Puffery Settings</i>
Weak Group Setting			
Average Audit Index	82.14	52.37	29.77
Average Fraud Index	121.90	137.08	-15.18
Average Payoff Differences (Manager Payoff minus Auditor Payoffs)	32.31	45.08	-12.77
Strong Group Setting			
Audit Index	118.69	90.01	28.68
Fraud Index	93.43	126.58	-33.15
Average Payoff Differences (Manager Payoff minus Auditor Payoffs)	21.95	32.56	10.61
Differences of Measures across the Group Settings			
Audit Index Comparison	-36.54	-37.64	
Fraud Index Comparison	28.47	10.50	
Average Payoff Differences Comparison	10.36	12.52	

^a n = 11 in each of the four settings: NP/WG, NP/SG, YP/WG, and YP/SG. The audit index is calculated by assigning the value of 0 to the trust audit, 1 to the Nash audit, and 2 to the defensive audit. These values are averaged and multiplied by 100. Thus, a lower audit index score indicates the auditor is more trusting. I calculated the manager's fraud index by assigning a value of 0 to the cooperate fraud level, 1 to the Nash fraud level, and 2 to the cheat fraud level. These values are averaged and multiplied by 100. Thus, a lower fraud index score indicates the manager is more cooperative. The payoff difference measure is calculated by subtracting the auditor's total expected payoffs from the manager's expected payoff.

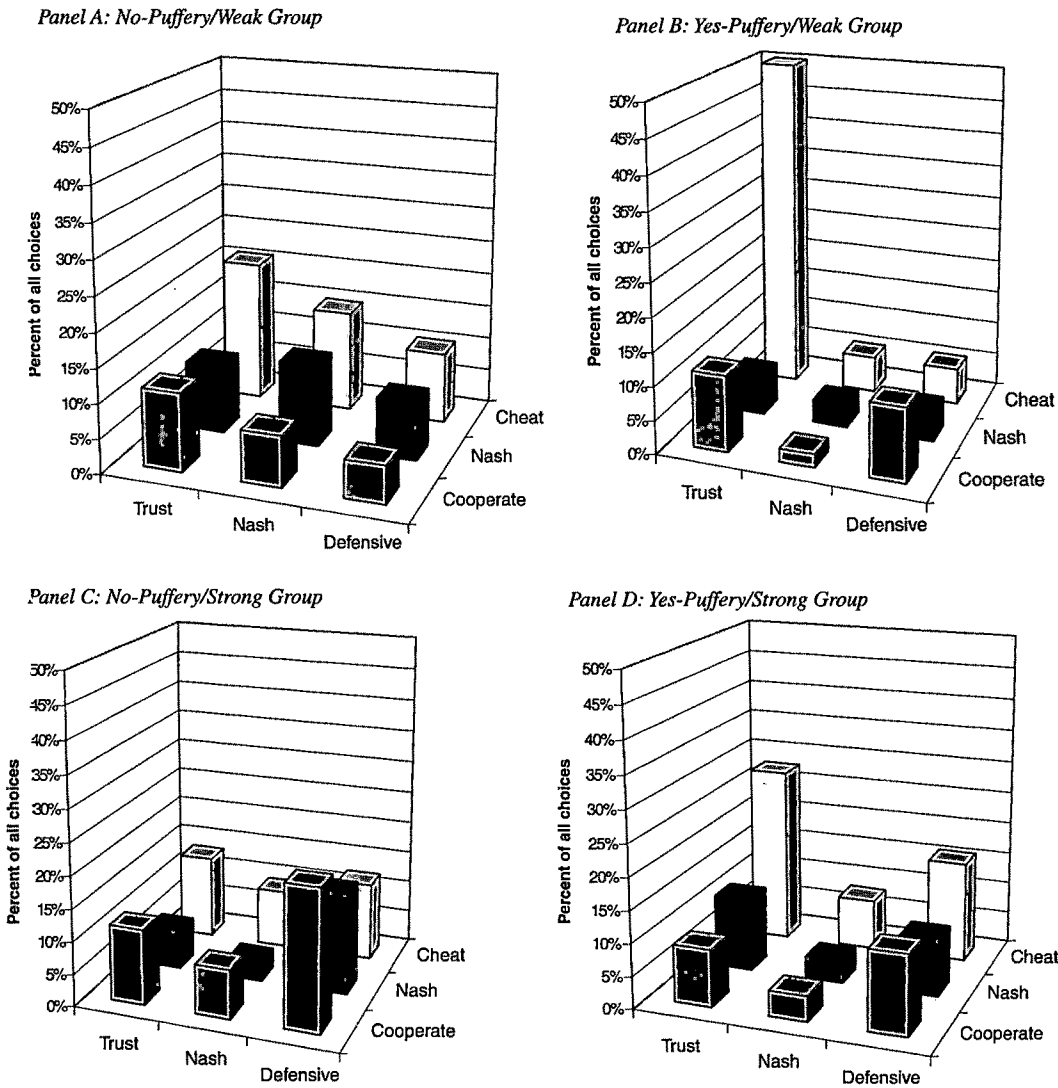
Thus, the data do not support either H1 (Nash) or H1 (Trust). These baseline findings serve as a benchmark for testing the incremental influence of the puffery and group manipulations in H2 and H3.

Results for H2: Comparing the NP Settings to the YP Settings

In the YP settings, managers could form reputations for trustworthiness during the puffery-action phase by sending auditors a message indicating a willingness to cooperate and then selecting a low fraud level. Managers did this in every round of the puffery-action phase, except one, when no message was sent (details are available upon request).

Table 3 indicates that auditors are more trusting in the YP settings than in the NP settings and the ANOVA results in Panel A of Table 4 show that the puffery effect is significant ($p = 0.03$). Thus, managers' puffery induced additional trust on the part of auditors. Moreover, Panel B of Table 4 shows that auditors' higher levels of trust were not justified by managers' fraud levels. Specifically, there is no significant difference ($p = 0.15$) between managers' fraud level in the NP and YP settings. These results are consistent with H2 (SSB). Figure 2 depicts these results. The most striking finding shown in Figure 2 is

FIGURE 2
Audit and Fraud Choices



The managers' fraud choices include cooperate (lowest level of fraud), Nash (middle level of fraud), and cheat (highest level of fraud) levels. The auditors' audit choices as trust (lowest audit), Nash (middle audit), and defensive (highest audit) levels.

the increased frequency of the trust/cheat combination in the YP/WG setting (shown in Panel B), relative to the NP/WG setting (shown in Panel A).

Panel C of Table 4 compares average payoff differences using ANOVA. These results, together with data from Table 3 show that the payoff difference was significantly higher in the YP setting than in the NP setting ($p = 0.013$). Thus, managers' puffery induced additional and unwarranted trust on the part of the auditors. As a result of falling prey to managers' puffery, auditors received lower average payoffs.

TABLE 4
Analysis of Variance for the Regular Play Phase^a

Panel A: Dependent Variable: Audit Index

<i>Effect</i>	<i>df</i>	<i>Sum of Squares</i>	<i>F-statistic</i>	<i>p-value</i>
Puffery	1	9,396	4.85	0.03
Group	1	15,136	7.83	0.01
Puffery × Group	1	3.3	0.00	0.96
Error	40	77,282		

Panel B: Dependent Variable: Fraud Index

<i>Effect</i>	<i>df</i>	<i>Sum of Squares</i>	<i>F-statistic</i>	<i>p-value</i>
Puffery	1	6,422	2.1	0.15
Group	1	4,176	1.4	0.25
Puffery × Group	1	888	0.3	0.59
Error	40	121,232		

Panel C: Dependent Variable: Payoff Difference

<i>Effect</i>	<i>df</i>	<i>Sum of Squares</i>	<i>F-statistic</i>	<i>p-value</i>
Puffery	1	551,618	6.7	0.01
Group	1	509,185	6.2	0.02
Puffery × Group	1	2,647	0.0	0.86
Error	40	3,288,742		

^a The audit index is calculated by assigning the value of 0 to the trust audit, 1 to the Nash audit, and 2 to the defensive audit. These values are averaged and multiplied by 100. Thus, a lower audit index score indicates the auditor is more trusting. I calculated the manager's fraud index by assigning a value of 0 to the cooperate fraud level, 1 to the Nash fraud level, and 2 to the cheat fraud level. These values are averaged and multiplied by 100. Thus, a lower fraud index score indicates the manager is more cooperative. The payoff difference measure is calculated by subtracting the auditor's total expected payoffs from the manager's expected payoff.

Results for H3: Comparing the WG and the SG settings

Hypothesis 3 investigates the extent of the SG treatment on auditors' choices. Figure 2, Panels C and D summarize audit and fraud choices. Notice the frequency of the trust/cheat combination appears somewhat higher in the YP/SG setting (shown in Panel D) than in the NP/SG setting (shown in Panel C). This finding is reinforced in Table 3, which shows that auditors are less trusting in the SG setting than in the WG settings, and the ANOVA results from Panel A of Table 4 show that this difference is significant ($p = 0.01$). Thus, the group treatment prompts auditors to adopt a defensive posture. Table 4, Panel B shows that there is no significant difference in the managers' fraud index across the group settings ($p = 0.25$), indicating that the lower level of auditors' trust was not triggered by a higher level of managerial fraud. This result supports H3, that the group treatment affects auditor choices.

Finally, Table 4, Panel C and results from Table 3 indicate that the payoff differences are lower in the SG setting ($p = 0.02$). Thus, the SG treatment led auditors to choose higher

audit levels without managers selecting higher fraud levels, leading to a smaller payoff disparity between auditors and managers. Panels A, B, and C of Table 4 also indicate that there are no significant interactions between the puffery and group effects for any of the three measures. Thus, in my setting, the puffery effects and the group effects offset one another, such that group affiliations appear to ameliorate the dysfunctional tendency to fall prey to puffery.

Results for H4: Comparing Audit Levels across the Resolution and Regular Play Phases

Hypothesis 4 (Null) predicts auditors' choices will not change after observing managers' choices from the regular play phase. The alternative predicts that information auditors receive about the managers' fraud level will lead them to choose higher levels of auditing during the resolution phase. Evidence that auditors increase their level of auditing during the resolution phase, would suggest that auditors' expectations are not in equilibrium in the regular play phase. Table 5 summarizes the three measures for the regular play and the resolution phases. For NP/WG, the audit index was higher in the resolution phase (104.55), than in the regular play rounds (82.14). This difference is marginally significant ($p = 0.07$) based on 11 matched pairs. There is a marginally significant increase in auditing in the other three settings, based on the average audit index. When I pool all the data across the 44 subjects, I do observe a significant increase in audit levels in the resolution phases ($p = 0.01$). This suggests that information about managers' choices affects auditors' strategies.

Analysis of the average fraud index (Panel B of Table 5) indicates that managers chose significantly less fraud in the resolution phase of the YP/WG setting, suggesting that managers were aware that they used puffery to condition auditors' expectations during the regular play phase, and expected retribution during the resolution phase. In addition, the average payoff differences (Panel C of Table 5) are significantly higher in the regular play than in the resolution phases for the NP/WG, YP/WG, and YP/SG settings. That is, auditors audited more and the managers chose less fraud in the resolution phase, so the auditors had a lower payoff disadvantage in the resolution phase.

VI. DISCUSSION

For independent auditors to enhance the credibility of financial statements, auditors must represent the interests of external parties, not those of management. For the past 30 years, policymakers have debated the extent to which economic forces impair auditors' ability to provide high-quality audits. Prior research has studied several features of the market for audits to determine their effects on audit quality: (1) low-balling—setting audit fees in early periods below the cost of the audit (see Schatzberg 1990; Dye 1991; Craswell and Francis 1999); (2) providing management advisory services (Antle et al. 1997); and (3) direct and indirect financial and family ties (SEC 2000). Research has argued that counterbalancing forces reduce auditors' incentives to misreport, including: (1) legal sanctions imposed by courts/regulators, and (2) market penalties from loss of reputational capital. Thus, auditors face trade-offs between economic rents from misreporting vs. potential economic penalties for misreporting. The policy debate typically centers on the strength of these various forces (U.S. Government Accounting Office 1996; Public Oversight Board Advisory Panel on Auditor Independence 1994; SEC 2000).

However, psychological as well as economic forces can impair audit quality. Bazerman et al. (1997) argue that even honest auditors can fall prey to self-serving biases that prevent them from conducting impartial audits. My research investigates self-serving bias in an

TABLE 5
Average Audit Index, Average Fraud Index, and Average Payoff Difference for Regular Play and Resolution Phases^a

Panel A: Average Audit Index

	<u>No Puffery/ Weak Group</u>	<u>No Puffery/ Strong Group</u>	<u>Yes Puffery/ Weak Group</u>	<u>Yes Puffery/ Strong Group</u>
Average Audit Index for:				
Regular Play Rounds	82.14	118.69	52.37	90.01
Resolution Play Rounds	104.55	127.27	90.91	121.21
Difference	-22.41	-8.58	-38.54	-31.20
	p = 0.07 ^b	p = 0.34	p = 0.09	p = 0.10

Panel B: Average Fraud Index

	<u>No Puffery/ Weak Group</u>	<u>No Puffery/ Strong Group</u>	<u>Yes Puffery/ Weak Group</u>	<u>Yes Puffery/ Strong Group</u>
Average Fraud Index for:				
Regular Play Rounds	121.90	93.43	137.08	126.58
Resolution Play Rounds	96.96	121.21	72.72	93.93
Difference	24.94	-27.78	64.36	32.65
	p = 0.13	p = 0.12	p = 0.02	p = 0.11

Panel C: Average Payoff Differences

	<u>No Puffery/ Weak Group</u>	<u>No Puffery/ Strong Group</u>	<u>Yes Puffery/ Weak Group</u>	<u>Yes Puffery/ Strong Group</u>
Average Payoff Difference for				
Regular Play Rounds	32.31	21.95	45.08	32.56
Average Payoff Difference for				
Resolution Play Rounds	24.89	22.84	25.47	20.89
Effect of Resolution Phase	7.42	0.89	19.61	11.67
	p = 0.05	p = 0.45	p = 0.02	p = 0.06

^a n = 11 in each of the four settings: no puffery/weak group; no puffery/strong group; yes puffery/weak group; yes puffery/strong group. Audit, fraud, and payoff indexes are defined in Table 3.

^b The p-values are based on two-tailed approximate randomization tests.

experimental environment that follows Kachelmeier's (1996a) suggestion to incorporate both economic and psychological forces. Auditors do exhibit self-serving bias in my setting; however, auditor group affiliation neutralizes this bias.

My research makes two primary contributions. First, it challenges the Bazerman et al. (1997) conclusion that auditors cannot be independent because they fall prey to self-serving biases. Although my results suggest that self-serving bias can arise, noneconomic forces such as group affiliations can help counter such biases. One implication is that accounting firms should continue to create and maintain strong group cohesion among audit teams, and to encourage auditors to affiliate with professional organizations.

A second contribution is the development of an experimental setting that incorporates key (stylized) features of the audit environment: (1) auditors benefit from trustworthy clients, but auditors may not be able to determine the client's true trustworthiness in the short

run, and (2) asymmetric incentives and information between the manager and the auditor create ambiguity about the other player's incentives. This is a parsimonious setting for investigating audit-quality issues with both economic and noneconomic components.

Like all research, this study is subject to several limitations. First, I used specific functions and parameter values to operationalize the theory for my experimental setting, and it is not clear how sensitive the findings are to these selected values. Second, the setting has a game-theoretic basis, and the results may not generalize to market or organizational settings. Third, I used student-subjects for the experiment rather than experienced auditors and managers. This may not be a material shortcoming for two reasons. First, previous research concludes that student-subjects may be most appropriate in experiments that do not incorporate mundane realism (Ball and Cech 1996). Second, Bazerman et al.'s (1997) conclusion was based on research that used student-subjects, and it seems reasonable to use the same subject pool to investigate the robustness of their conclusions. Another potential limitation of the research is that the experiment is devoid of context (i.e., without mention of fraud, audit, or self-serving bias). As Haynes and Kachelmeier (1998) suggest, context can affect subject behavior in important ways.

Future research could modify this setting to incorporate a more realistic context. One possibility is to build on Kadous et al. (2001) to investigate how motivated reasoning might affect auditor-subjects' strategies, especially when there are competing goals from the client- and the auditor-group members. A second avenue for future research is to investigate self-serving bias when client selection and group formation are endogenous.

REFERENCES

- Abrams, D., and M. Hogg. 1999. *Social Identity and Social Cognition*. Oxford, U.K.: Blackwell Publishers.
- Antle, R., P. Griffin, D. Teece, and O. Williamson. 1997. An economic analysis of auditor independence for a multi-client, multi-service public accounting firm. In *Serving the Public Interest: A New Conceptual Framework for Auditor Independence*. A report prepared on behalf of the AICPA in connection with the presentation to the Independence Standards Board. Available at: <http://ftp.aicpa.org/public/download/members/div/secps/isb/0117194.doc>.
- Babcock, L., and G. Loewenstein. 1997. Explaining bargaining impasse: The role of self-serving bias. *Journal of Economic Perspectives* 11 (Winter): 109–126.
- Ball, S., and P. Cech. 1996. Subject pool choice and treatment effects in economic laboratory research. *Research in Experimental Economics* 6: 239–292.
- Bazerman, M., K. Morgan, and G. Loewenstein. 1997. The impossibility of auditor independence. *Sloan Management Review* (Summer): 89–94.
- Berg, J., J. Dickhaut, and K. McCabe. 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10 (July): 122–142.
- Bloomfield, R. 1995. Strategic dependence and inherent risk assessments. *The Accounting Review* 70 (January): 71–90.
- . 1997. Strategic dependence and the assessment of fraud risk: A laboratory study. *The Accounting Review* 72 (October): 517–538.
- Cooper, R., D. DeJong, R. Forsythe, and T. Ross. 1989. Communication in the battle of the sexes: Some experimental results. *Rand Journal of Economics* 20 (Winter): 568–587.
- Craswell, A., and J. Francis. 1999. Pricing initial audit engagements: A test of competing theories. *The Accounting Review* 74 (April): 201–216.
- Dawes, R., and R. Thaler. 1988. Anomalies: Cooperation. *Journal of Economic Perspectives* 2 (Summer): 187–197.
- Dye, R. 1991. Informationally motivated auditor replacement. *Journal of Accounting and Economics* 14 (December): 347–374.
- Edgington, E. 1987. *Randomization Tests*. 2nd edition. New York, NY: Marcel Dekker.

- Haynes, C., and S. Kachelmeier. 1998. The effects of accounting contexts on accounting decisions: A synthesis of cognitive and economic perspectives in accounting experimentation. *Journal of Accounting Literature* 17: 97–136.
- Kachelmeier, S. 1991. A laboratory market investigation of the demand for strategic auditing. *AUDITING: A Journal of Practice & Theory* 10 (Supplement): 25–48.
- . 1996a. Discussion of: Tax advice and reporting under uncertainty: Theory and experimental evidence. *Contemporary Accounting Research* 13 (Spring): 81–89.
- . 1996b. Do cosmetic reporting variations affect market behavior? A laboratory study of the accounting emphasis on unavoidable costs. *Review of Accounting Studies* 1 (2): 115–140.
- , and M. Shehata. 1997. Internal auditing and voluntary cooperation in firms: Cross cultural experiment. *The Accounting Review* 72 (July): 407–431.
- Kadous, K., J. Kennedy, and M. Peecher. 2001. Can accuracy goals reduce auditor objectivity? Working paper, University of Washington and University of Illinois.
- King, R., and R. Schwartz. 2000. An experimental investigation of auditors' liability: Implications for social welfare and explorations of deviations from theoretical predictions. *The Accounting Review* (October): 429–451.
- Libby, R., R. Bloomfield, and M. Nelson. 2001. Experimental research in financial accounting. *Accounting, Organizations, and Society* (forthcoming).
- Matsumura, E., and R. Tucker. 1992. Fraud detection: A theoretical foundation. *The Accounting Review* 67 (October): 753–782.
- Mayhew, B., J. Schatzberg, and G. Sevcik. 2000. Entrepreneur choice of auditor and retained ownership in IPO markets: Experimental evidence. Working paper, University of Wisconsin–Madison.
- Palmrose, Z-V. 1994. The joint and several versus proportionate liability debate: An empirical investigation of audit-related litigation. *Stanford Journal of Law, Business & Finance* 1 (Fall): 53–72.
- Public Oversight Board Advisory Panel on Auditor Independence. 1994. *Strengthening the Professionalism of the Independent Auditor*. September. Stamford, CT: POB.
- Roth, A. 1987. *Laboratory Experimentation in Economics*. Cambridge, U.K.: Cambridge University Press.
- Schatzberg, J. 1990. A laboratory investigation of low-balling in audit pricing. *The Accounting Review* (April): 337–362.
- Securities and Exchange Commission (SEC). 2000. *Hearing on Auditor Independence*. Wednesday, July 26, 2000. Available at: <http://www.sec.gov/rules/extra/audmin.htm>.
- U. S. General Accounting Office (GAO). 1996. *The Accounting Profession—Major Issues: Progress and Concerns*. Report to the Ranking Minority Member, Committee on Commerce, House of Representatives: September. Washington, D.C.: Government Printing Office.

Do Conference Calls Affect Analysts' Forecasts?

Robert M. Bowen

University of Washington

Angela K. Davis

Washington University in St. Louis

Dawn A. Matsumoto

University of Washington

ABSTRACT: In 1998, the SEC expressed concern that conference calls encourage selective disclosure by revealing new information to financial analysts privy to the call. This study investigates whether the regular use of earnings-related conference calls increases the amount of information available to financial analysts by examining the effect of conference calls on analysts' forecast error and dispersion. Results indicate that conference calls increase analysts' ability to forecast earnings accurately, suggesting that these calls increase the total information available about a firm. We also find some evidence that conference calls decrease dispersion among analysts. Given conference calls were generally restricted during our sample period, our evidence suggests that conference calls may have contributed to an information gap between analysts privy to the call and the remainder of the investment community.

We also investigate whether conference calls differentially affect analysts' forecast errors depending on analysts' prior forecasting ability or brokerage-house affiliation. We find evidence suggesting that analysts with relatively weak prior forecasting performance benefit more from conference calls, suggesting that conference calls help "level the playing field" across analysts.

Keywords: *conference calls; security analysts; forecast error; forecast dispersion; disclosure; information environment; Reg FD.*

Data Availability: *Data are commercially available from the sources identified in the text.*

We thank Dave Burgstahler, Brian Bushee, Larry DuCharme, Amy Hutton, Marilyn Johnson, Bjorn Jorgenson, Jane Kennedy, Greg Miller, James Myers, Linda Myers, Chris Noe, Shiva Rajgopal, Jean Schlinger, Terry Shevlin, D. Shores, and an anonymous referee for helpful comments. The comments of workshop participants at Harvard University, the University of New Hampshire, and the University of Washington are greatly appreciated. We are grateful to First Call Corporation for providing data used in this study.

*Submitted March 2000
Accepted September 2001*

I. INTRODUCTION

Conference calls are a common method that senior management uses to communicate with financial analysts. A 1998 study by the National Investor Relations Institute (NIRI 1998a) reports that, of the member firms surveyed, 82 percent use conference calls—up from 61 percent reported in a similar survey conducted in 1995. Despite the increasing popularity of conference calls, there is little empirical evidence on the effects of these calls on their primary audience—financial analysts. This study examines the effects of conference calls on analysts' earnings forecasts before the Securities and Exchange Commission (SEC) issued Regulation Fair Disclosure (S7-31-19, available at: <http://www.sec.gov/rules/proposed/34-42259.htm>) in August 2000 (hereafter Reg FD).¹

During the 1998 "SEC Speaks" Conference, SEC Chairman Arthur Levitt expressed concern that analyst calls (including conference calls) represent *increased disclosure* and therefore put the general public at a competitive disadvantage in the search for information. Whether conference calls represent *increased disclosure* or simply a *new method* of disclosure is an empirical question. *Increased disclosure* means that the firm releases new, material information (or that call participants infer new information from the interaction with management during the call). Alternatively, conference calls may represent merely a *new method* of disclosure in that the same information released in a conference call is available through other sources (e.g., individual discussions, press releases), although perhaps with different timing.²

Increased disclosure leads to *selective disclosure* when new information is available to some, but not all, capital market participants. To the extent that analysts were the primary audience for conference calls, new information released during these calls could represent selective disclosure.³ Thus, we investigate whether conference calls reveal new information—one necessary condition for selective disclosure to occur.

Concern about selective disclosure influenced the SEC to pass Reg FD, which requires that (1) when an issuer intentionally discloses material information, it does so through public disclosure, not through selective disclosure; and (2) whenever an issuer learns that it has made a non-intentional material selective disclosure, the issuer makes prompt public disclosure of that information. Our investigation of conference calls as one potential source of selective disclosure contributes to the ongoing controversy surrounding the passage of Reg FD.

We evaluate whether conference calls represent increased disclosure by examining their effect on two properties of analysts' forecasts—forecast error and forecast dispersion. If regular use of earning-related conference calls increases disclosure (i.e., management discloses additional information during conference calls that is not otherwise disclosed to

¹ Although conference calls were originally directed at financial analysts and large investors, the increasing popularity of equity investing in the mid- to late-1990s likely increased the demand among individual investors to listen to these calls. For example, in March 1999, Bestcalls.com launched a web site publicizing the dates and times of conference calls open to individual investors. However, firms still maintained control over call participants and many elected not to allow individuals access to their calls. Around the same time, some firms began live broadcasts of their conference calls via the Internet (webcasts). These calls are obviously open to all listeners, including the financial press. Given the 1995–1998 sample period used in this study, individuals were unlikely to be allowed access to the conference calls examined in this study. Bushee et al. (2001) examine a sample of conference calls open to individual investors and investigate the determinants and effects of allowing unlimited access to conference calls either via telephone or webcast.

² For example, Lou Thompson, president of NIRI, argued that conference calls simply provide perspective and explanation of the fundamental news previously released in an earnings announcement (NIRI 1998b).

³ A recent NIRI study reports that, of firms hosting conference calls, 82 percent allow individual investors real-time access vs. 27 percent in 1998, and 74 percent allow the financial press access vs. only 14 percent in 1998 (NIRI 2000).

analysts), then conference calls should lead to lower forecast error and forecast dispersion in the period around the call. For our cross-sectional tests, we compute the change in forecast error and the change in forecast dispersion during a quarter, and compare the change in error and the change in dispersion between firms hosting conference calls that quarter and firms not hosting conference calls. For our within-firm tests, we compare the change in forecast error and the change in dispersion before and after each firm begins to regularly host conference calls.

Cross-sectional tests reveal that the average error in analysts' forecasts declines more in conference call quarters than in non-conference call quarters. Further, the relatively greater decrease in error for conference call quarters persists through the next earnings announcement date, suggesting a sustained information advantage. These results are robust to a number of controls and sensitivity checks. Our within-firm tests provide similar, though somewhat weaker, results. For a subsample of firms that regularly host conference calls, forecast error declines more in quarters after the firm begins hosting conference calls than in earlier quarters before the firm used conference calls.

Similar to our findings for forecast error, our cross-sectional tests on forecast dispersion indicate that dispersion in analysts' forecasts tends to decrease during the quarter for all firms, but that the decrease in dispersion is *relatively* greater in conference call quarters than in non-conference call quarters. Also, this relatively greater decrease in dispersion persists through the next earnings announcement date—again consistent with a sustained effect on the information environment. However, the results of our within-firm tests are not statistically significant.

We next investigate whether conference calls “level the playing field” across analysts by testing whether the decrease in forecast error associated with conference calls differs across analyst types. Analysts with superior forecasting ability or affiliated with more prestigious brokerage firms may receive differential benefits from conference calls than analysts with relatively less ability or less prestigious affiliation. We compare the change in error for these different types of analysts and examine the effect of conference calls across types. We find that both types of analysts benefit from conference calls, but analysts with relatively weaker prior forecasting ability appear to benefit more from conference calls.

Given the controversy surrounding Reg FD, evidence on the effect of conference calls on capital market participants is important from a policy perspective. Our evidence suggests that conference calls increase the information available to analysts. Given that firms generally restricted public access to conference calls during our sample period, our results suggest that conference calls likely contributed to an information gap between analysts privy to the call and the remainder of the investment community, consistent with the SEC's concerns.

From an academic perspective, our study provides new evidence on the effects of disclosure on analysts' forecasts. Lang and Lundholm (1996) find that higher Financial Analyst Federation ratings of firms' disclosure policies are associated with more accuracy and less dispersion in analysts' forecasts. In contrast, Healy et al. (1999) do not find a significant relation between forecast dispersion and sustained increases in firms' disclosure policies (measured by analysts' ratings of firms' disclosures from the Association of Investment Management and Research). Our study differs from Lang and Lundholm (1996) and Healy et al. (1999) in that we are better able to isolate the (potential) disclosure event and measure more precisely the effect of the conference call event on analysts' forecasts. Two other studies focus more precisely on the effects of specific disclosure events on analysts' forecasts. Barron et al. (1999) find that high-quality disclosures in the management discussion and analysis (MD&A) section of the 10-K are associated with more accurate

forecasts and lower dispersion; however, Francis et al. (1997) find that managers' presentations to security analysts have no effect on analysts' forecast error or dispersion. These seemingly conflicting results suggest that not all disclosure events or measures have similar effects on analysts' forecasts. Our results suggest that conference calls are an important disclosure event, at least before the passage of Reg FD.

In the next section, we develop empirical predictions. Section III discusses sample selection criteria and the research design. Section IV presents empirical results on the effect of conference calls on forecast error and dispersion. Section V presents similar results for different types of analysts. The last section offers concluding remarks.

II. DEVELOPMENT OF EMPIRICAL PREDICTIONS

The fact that analysts participate in conference calls suggests that calls provide material information that is potentially useful for forecasting future earnings and making stock recommendations. However, it is possible that conference calls simply expand on previously released news or partially replace other mechanisms for disclosure (e.g., one-on-one conversations with analysts).⁴ Thus, empirical evidence is necessary to assess whether conference calls provide new information.

Prior evidence suggests that managers disclose at least some information during conference calls. Frankel et al. (1999) find unusually high levels of return volatility and trading volume during conference call periods. Their findings provide indirect evidence suggesting that conference calls introduce new, material information to the analyst community.⁵ Further, SEC concerns about selective disclosure during our sample period and the subsequent adoption of Reg FD, suggest that managers may have released information during conference calls. Thus, given prior academic evidence and regulatory concerns, we predict that conference calls result in the release of material new information.

To generate specific empirical predictions regarding the effect of conference calls on analysts' forecasts, we use the framework developed by Barron et al. (1998) (hereafter BKLS), which expresses two observable properties of analysts' forecasts—error in the mean consensus forecast and dispersion in individual forecasts—as functions of the amount or “precision” of analysts' public and private information.⁶ In the BKLS framework, analysts receive information about earnings—both public information that is common across all analysts and private information that is idiosyncratic and uncorrelated across analysts. If firms provide additional information during conference calls, then conference calls increase the precision of public information.

Although it is intuitive that more precise public information should decrease error in the mean forecast, this prediction is not directly implied in the BKLS framework. To the extent that more precise public information reduces uncertainties about the firm's future prospects, increased precision in public information decreases forecast error. However, if the increased precision in public information also leads analysts to rely less on private information, then forecast error could actually increase (because the errors in private signals are uncorrelated by definition and, therefore, tend to cancel out when averaged across many

⁴ Anecdotal reports suggest a significant amount of one-on-one phone interaction exists between analysts and managers (McCafferty 1997). In a survey conducted prior to 1981, analysts ranked “interviews with company executives” as the most important source of information about specific companies, and 65 percent of the companies surveyed provided some assistance to security analysts (Lees 1981). Thus, it is possible that firms not using conference calls provide the same information to analysts via other means such as press releases and one-on-one phone conversations.

⁵ Frankel et al. (1999) recognize that their tests establish primarily a timing advantage associated with conference calls, and do not necessarily demonstrate a sustained increase in information.

⁶ The Appendix provides a detailed discussion of the BKLS framework and implications for our predictions.

analysts). Thus, the extent to which increased precision in public information increases or decreases the error in the mean forecast depends on the precision of public relative to private information, and the number of individual analysts contributing to the mean consensus forecast. For reasons discussed in the Appendix, we argue that the effect of decreased uncertainty dominates and, thus, if conference calls increase the precision of public information, then we expect these calls to be associated with decreased error in analysts' earnings forecasts.⁷ Our first empirical prediction (P1) is as follows:

P1: Forecast error will decrease more during conference call quarters than during non-conference call quarters.

The effect of increased precision on forecast dispersion is more straightforward in the BKLS framework—increased precision in public information decreases the dispersion in analysts' forecasts. Thus, if conference calls increase the precision of public information, then these calls should also be associated with decreased dispersion, leading to our second empirical prediction:

P2: Forecast dispersion will decrease more during conference call quarters than during non-conference call quarters.

III. RESEARCH DESIGN

Dependent Variables

The general form of each dependent variable is the scaled change in the forecast characteristic—error (ERROR) or dispersion (DISP)—before and after a conference call:

$$\frac{\text{post-conference call measure} - \text{pre-conference call measure}}{\text{stock price at pre-conference call date}}$$

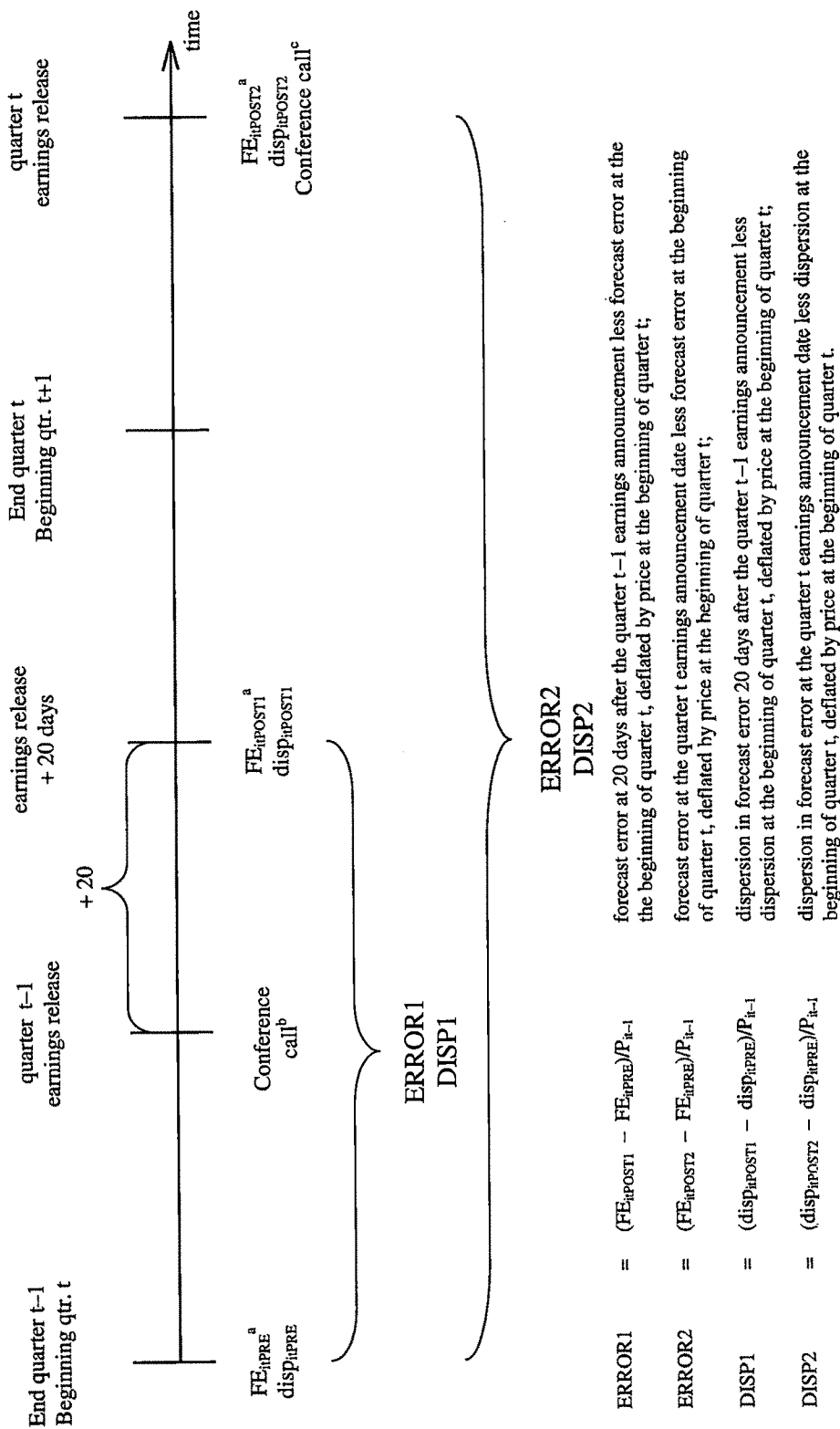
We measure forecast error for each firm-quarter as the absolute value of the difference between the mean of analysts' estimates and actual earnings per share for a given quarter.⁸ We measure dispersion for each firm-quarter as the standard deviation of analysts' individual forecast estimates. Post vs. pre differences isolate the effects of the conference call and control for differences in the *levels* of forecast error and dispersion across firms. If forecast error decreases following conference calls, then the post-call forecast error will be less than the pre-call forecast error and the scaled change in forecast error (ERROR) will be negative. Similarly, if forecast dispersion decreases following conference calls, then the scaled change in forecast dispersion (DISP) will be negative. Figure 1 illustrates the timeline of events and describes the measurement of the dependent variables in detail.

The *pre-conference call* component of the dependent variables is the forecast error or forecast dispersion for quarter *t* measured at the end of quarter *t*–1 (i.e., beginning of

⁷ We expect the effect of increased precision of public information to outweigh the effect of reduced diversification because we argue that public information is on average more precise than private information. Disclosures provided by the firm such as annual reports, quarterly filings, and earnings announcements, are examples of important public information. In addition, the BKLS framework would consider any *common* information disclosed privately to all analysts to be effectively public information.

⁸ In the BKLS framework, error in the mean forecast is defined as the squared difference between the mean analyst forecast and actual earnings. We use the absolute value of the difference rather than squaring the differences because squaring places greater weight on large forecast errors. However, our main results hold using squared differences.

FIGURE 1
Timeline of Events and Measurement of the Dependent Variables



(Continued on next page)

FIGURE 1 (Continued)

^a $FE_{it,h} = |F_{it,h} - E_{it,h}|$ the absolute value of the consensus analyst forecast error for firm i for quarter t at horizon h (i.e., PRE, POST1, POST2). $F_{it,h}$ is the mean consensus analyst forecast for firm i for quarter t at horizon h as calculated using Zack's individual analyst estimate history. E_{it} is the actual earnings for quarter t as reported on Zack's earnings surprise file; and

$disp_{it,h}$ = standard deviation of analysts' forecasts of firm i 's earnings for quarter t at horizon h (i.e., PRE, POST1, POST2).

Thus,

$FE_{it,PRE}$ = consensus forecast error for firm i for quarter t as of the beginning of the quarter;

$disp_{it,PRE}$ = standard deviation of analysts' forecasts of firm i 's earnings for quarter t as of the beginning of the quarter;

$FE_{it,POST1}$ = consensus forecast error for firm i as of 20 days after quarter $t-1$'s earnings announcement;

$disp_{it,POST1}$ = standard deviation of analysts' forecasts of firm i 's earnings as of 20 days after quarter $t-1$'s earnings announcement;

$FE_{it,POST2}$ = consensus forecast error for firm i for quarter t earnings at the quarter t earnings announcement; and

$disp_{it,POST2}$ = standard deviation of analysts' forecasts of firm i 's earnings for quarter t as of quarter t 's earnings announcement.

^b Management discusses quarter $t-1$'s earnings release and answers analysts' questions.

^c Management discusses quarter t 's earnings release and answers analysts' questions.

quarter t). We consider quarter t a conference call quarter if the firm hosts a conference call relating to the announcement of quarter $t-1$'s earnings. Prior research (and evidence in Table 1 discussed below) suggests that firms typically host conference calls following a quarterly earnings release (Tasker 1998; Frankel et al. 1999). During this conference call, management discusses earnings for the prior quarter (quarter $t-1$) and provides analysts with information potentially useful in forecasting earnings for the upcoming quarter (quarter t). While the conference call content may also help analysts forecast future quarters' earnings (in $t+1$, $t+2$, etc.), the conference call information likely has the greatest influence on analysts' forecasts of quarter t 's earnings.

We measure *post-conference call* forecast error or forecast dispersion at two alternative points in time, resulting in two dependent variables for each prediction. As Figure 1 depicts, we take the first post-call measures (ERROR1 and DISP1) 20 calendar days after the quarter $t-1$ earnings announcement. This measure forms a relatively narrow window around conference calls, and we intend it to isolate the effects of the calls by capturing differences in information dissemination between conference call and non-conference call firms. These effects may be temporary, however, if non-conference call firms "catch up" by providing equivalent information after the 20-day window but before the next earnings announcement. To evaluate whether conference calls cause a longer-term difference in error and/or dispersion, we need to allow time for non-conference call firms to disclose information relevant to quarter t earnings. Thus, our second post-call measures (ERROR2 and DISP2) start at the same time but end later, upon the announcement of quarter t earnings.

For each of the three measurement dates depicted in Figure 1 (i.e., the pre-conference call and the two post-conference call dates), we use the most current outstanding forecast from each individual analyst—as long as it is within 90 days before the measurement date.⁹ The 90-day requirement reduces the effects of potentially stale forecasts, and includes only analysts who actively follow a firm. We then match each analyst's pre- and post-conference call forecasts and retain only those forecasts made by analysts with forecasts at all three horizons.

Conference Call Data and Sample Selection

First Call, an information provider to institutional investors, maintains a data set of daily conference call schedules, which we use to identify firms hosting conference calls. Table 1 breaks down the number of conference calls on the First Call database by year. The first three columns present all observations listed on the First Call database (excluding duplicate observations, missing data, and clearly erroneous conference call dates).¹⁰ The data indicate that firms' use of conference calls has grown steadily from 1995 (the database ends in April 1998, resulting in fewer observations in 1998). Approximately 3,800 firms have at least one conference call on the database.

⁹ The 90-day horizon means that some forecasts included in the first post-conference call measures were actually made *before* the conference calls. We include these forecasts because an analyst's failure to revise a forecast after an earnings announcement/conference call is indicative of the (lack of) information released. Our analysis does, however, control for the average age of the forecasts in the cross-sectional tests.

¹⁰ To evaluate the completeness and accuracy of the First Call data, Tasker (1998) obtained independent verification of conference call firms by contacting sell-side analysts and asking them to identify firms that held quarterly conference calls from a sample of firms in the industry the analyst followed. Tasker (1998) finds that the First Call data include 95 percent of the firms that regularly host conference calls. In addition, Frankel et al. (1999) select a sample of 188 firms listed on Compustat that the First Call database did not list as hosting conference calls. The authors contacted these firms individually and 92 percent confirm that they did not hold conference calls during 1995 (the year of their study). To the extent that the inaccuracies in the First Call database result in misclassifications of conference call and non-conference call quarters, the power of our tests is reduced.

TABLE 1
Descriptive Data on Conference Calls

Year	Total Conference Calls on First Call Database ^a			Conference Calls Near Earnings Announcement Date ^b			Conference Calls with Forecast Data Available on Zack's Individual Tape ^c		
	Calls	Firms	Calls/Firm	Calls	Firms	Calls/Firm	Calls	Firms	Calls/Firm
1995	3,020	1,488	2.03	2,125	1,209	1.76	1,513	872	1.74
1996	6,011	2,482	2.42	4,456	2,149	2.07	3,120	1,535	2.03
1997	8,008	3,110	2.57	5,873	2,725	2.16	4,283	2,131	2.01
1998	2,507	1,742	1.44	1,861	1,493	1.25	819	803	1.02
Total	19,546	8,822	2.22	14,315	7,576	1.89	9,735	5,341	1.82
Total unique sample firms		3,836			3,366			2,607	

^a The complete data set from First Call Corporation included 20,533 conference calls. 619 were duplicates (same firm and date of conference call); 20 observations listed dates for the conference call beyond 1998; and 348 observations had missing data (for either firm identification or date of conference call).

^b We classify conference calls as near an earnings announcement if the date of the conference call falls between three days prior to and five days after the earnings announcement date listed on Zack's surprise file. We count multiple conference calls held within this nine-day window only once (eliminating 638 conference calls).

^c Conference calls with at least one forecast made within the last 90 days from two dates corresponding to the pre-conference call and post-conference call date. See Table 2 for further sample selection criteria and Figure 1 for further description of pre- and post-conference call dates.

The next three columns of Table 1 list the number of conference calls hosted within a nine-day window (-3 to $+5$) surrounding the earnings announcement date (as listed on Zack's Investment Research's surprise file). Approximately 75 percent of all conference calls occur within this nine-day window, suggesting that most conference calls take place in conjunction with earnings announcements. The last three columns present similar data for firm-quarters with forecast data available on Zack's individual analyst estimate history.

Table 2 summarizes our sample selection process. The initial sample consists of firm-quarters with earnings announcement dates for both quarter t and $t-1$ available on Zack's earnings surprise history files for the period January 1995 through March 1998 (corresponding to the period covered by the First Call database). We merged the initial sample with the First Call data to identify firm-quarters where conference calls occur within a nine-day window (-3 to $+5$) surrounding the prior quarter's earnings announcement (13,253 firm-quarters).¹¹ We call these "conference call quarters." We classify firm-quarters in the initial sample that are not identified as conference call quarters as non-conference call quarters and use them as a control group (64,389 firm-quarters). From this combined sample (77,642 firm-quarters), we eliminate (1) firm-quarters where managers are "late" in announcing earnings (more than 45 [90] days past the end of the quarter for fiscal quarters 1-3 [4]); (2) firm-quarters that do not have at least two analysts with forecasts at both the pre- and post-conference call date; (3) firm-quarters without price data on Compustat; and (4) the top and bottom 1 percent of the dependent variable observations.¹²

Finally, three additional data selection criteria reduce the effects of potentially confounding factors:

- 1) We eliminate quarters where firms host conference calls between the fifth day after the quarter $t-1$ earnings announcement and the date of the quarter t earnings announcement.
- 2) For firms hosting at least one conference call, but that host conference calls in less than 50 percent of their firm-quarters after their first appearance on the First Call database ("sporadic conference call users"), we eliminate *all* firm-quarters from their first conference call onward, i.e., we include observations for these firms *only* before they begin their sporadic use of conference calls as non-conference call firm-quarters.¹³
- 3) For firms hosting conference calls in more than 50 percent of their firm-quarters after their first appearance on the First Call database (we call these "regular conference call users"), we eliminate all "non-conference call" quarters after the first conference call.

¹¹ Use of a narrower window (0 to $+2$) does not affect our results. Our tabulated results are based on the nine-day (-3 to $+5$) window because firms may differ in their timing of conference calls relative to the earnings release.

¹² The number of observations listed represents all firm-quarters with at least one dependent variable. Less than 2 percent of observations are eliminated since we determined the top and bottom 1 percent for each dependent variable separately. Eliminating the top and bottom 1 percent significantly affects the parametric tests reported in the next section (results are generally insignificant if we include these extreme observations). However, as might be expected, the nonparametric univariate tests are unaffected by the inclusion of these observations.

¹³ We define regular (vs. sporadic) users as firms that host conference calls in at least 50 percent of firm-quarters. We computed the percentage of quarters a firm hosts a conference call after the company's first appearance on the First Call database. Sixty-four percent of the firms with at least one conference call have a conference call percentage of greater than 50 percent. We test the sensitivity of our results to this sample selection criterion in Section IV.

TABLE 2
Sample Selection Criteria

	<u>Total</u>	<u>Conference Call Quarters^a</u>	<u>Non-Conference Call Quarters</u>
Quarters with earnings announcement dates for quarter t-1 available on Zack's surprise file (1995-1998)	90,173	14,315	75,858
No earnings announcement date for quarter t available on Zack's surprise file	<u>(12,531)</u>	<u>(1,062)</u>	<u>(11,469)</u>
Quarters with earnings announcement dates available on Zack's surprise file (1995-1998)	77,642	13,253	64,389
Earnings announcements not within 45 (90) days of quarter end for fiscal quarters 1-3 (4)	<u>(14,790)</u>	<u>(1,216)</u>	<u>(13,574)</u>
	62,852	12,037	50,815
No forecast revisions available on Zack's individual detail tape within 90 days of the PRE, POST1, and POST2 horizon dates (see Figure 1)	<u>(30,231)</u>	<u>(2,302)</u>	<u>(27,929)</u>
	32,621	9,735 ^b	22,886
Price data not available on quarterly Compustat	<u>(690)</u>	<u>(88)</u>	<u>(602)</u>
	31,931	9,647	22,284
Delete top and bottom 1 percent of dependent variable observations	<u>(343)</u>	<u>(78)</u>	<u>(265)</u>
	31,588	9,569	22,019
Delete firm-quarters with only one forecast	<u>(11,493)</u>	<u>(2,213)</u>	<u>(9,280)</u>
	20,095	7,356	12,739
Delete quarters with conference calls outside earnings announcement window	<u>(1,763)</u>	<u>(941)</u>	<u>(822)</u>
	18,332	6,415	11,917
Delete all quarters once conference calls begin for firms hosting at least one conference call but in 50 percent or less of firm quarters	<u>(4,338)</u>	<u>(1,690)</u>	<u>(2,648)</u>
	13,994	4,725	9,269
Delete non-conference call quarters once conference calls begin for firms with conference calls in greater than 50 percent of firm-quarters	<u>(1,439)</u>	<u>—</u>	<u>(1,439)</u>
Final sample	<u>12,555</u>	<u>4,725</u>	<u>7,830</u>

^a Quarters are classified as conference call quarters if the firm hosts a conference call within a nine-day window (-3 to +5) surrounding the prior quarter's earnings announcement.

^b Corresponds with the total conference calls with forecast data available from Zack's in Table 1.

The first criterion eliminates firm-quarters with "special" conference calls not directly related to earnings announcements. Conference calls that are unrelated to earnings announcements are more likely to result from some underlying news event (e.g., announcement of a merger), which could influence the change in forecast error or dispersion. Similarly, the second criterion eliminates all quarters once a firm starts hosting conference calls on a sporadic basis. In these quarters, firms are more likely to host conference calls that discuss unusual events than recurring earnings announcements. By focusing only on conference calls regularly held in conjunction with earnings announcements, we reduce the possibility that the choice to host a conference call is endogenously determined.¹⁴ Finally, since conference calls likely provide information useful for forecasting earnings beyond the next quarter, the third criterion eliminates subsequent non-conference call quarters for regular conference call users to avoid contamination from potential residual effects of previously held conference calls.

Control Variables

Our tests of the relation between conference calls and (1) changes in analyst forecast error and (2) changes in forecast dispersion, control for firm size, earnings surprise, and forecast age, which prior research has found to affect the *levels* of forecast error and dispersion (Lang and Lundholm 1996; Barron et al. 1998; Brown 2001). We define these variables as follows:

SIZE = log of the market value of the firm's equity at the beginning of quarter t ;
 SURP _{t} = $|\text{EPS}_t - \text{EPS}_{t-4}| / P_{t-4}$, where EPS _{t} is the primary earnings per share (including extraordinary items) for quarter t and P_{t-4} is the ending price per share at quarter $t-4$;

ΔAGE1 = average age of forecasts of quarter t earnings 20 days after quarter $t-1$ earnings announcement (POST1) less the average age of forecasts of quarter t earnings at the beginning of the quarter (PRE). Both averages are based on the forecasts used to compute the dependent variables, ERROR1 and DISP1; and

ΔAGE2 = average age of forecasts of quarter t at the earnings announcement date (POST2) less the average age of forecasts of quarter t earnings at the beginning of the quarter (PRE). Both averages are based on the forecasts used to compute the dependent variables, ERROR2 and DISP2.

SIZE proxies for the richness of the firm's information environment. Although a firm's information environment affects the *levels* of forecast error and dispersion, we do not expect it to affect the *changes* in forecast error and dispersion; however, we control for size as a precaution. Second, we use the change in quarterly earnings relative to the prior year to proxy for the difficulty in forecasting earnings. Following Lang and Lundholm (1996), we label this variable as "surprise" (SURP). We control for earning surprise in both the quarter being forecasted and the prior quarter since analysts' ability to forecast quarter t 's earnings likely depends on both the earnings surprise in the prior quarter (i.e., quarter $t-1$) and any information discussed during the conference call that suggests an upcoming earnings surprise. Third, since stale forecasts increase the levels of forecast error and forecast dispersion,

¹⁴ Results subsequently presented in Table 4 indicate that the pre-announcement levels of forecast error and dispersion are not higher in conference call quarters than in non-conference call quarters, suggesting that the decision to host conference calls is not endogenously driven by a manager's attempt to "guide" analysts' forecasts that are relatively inaccurate.

we control for differences in the average age of the consensus forecasts (ΔAGE). Brown (2001) finds forecast age is the most important determinant of forecast accuracy.

Finally, we add a new control variable—the pre-conference call levels of forecast error and dispersion. We use changes rather than levels of forecast error and dispersion to mitigate the effect of cross-sectional differences in information environments; however, the richness of a firm's information environment likely impacts the *initial* level of forecast error or dispersion, which could limit the potential for reductions in these measures. For example, firms with richer information environments likely have lower initial forecast errors and dispersion and, as a result, have less room for conference calls to further decrease error and dispersion. Thus, we include pre-conference call levels of forecast error and dispersion to control for any potential relation between the use of conference calls and the richness of a firm's information environment.¹⁵

Cross-Sectional Tests

The following four regressions form the basis of our cross-sectional tests:

$$\begin{aligned} \text{ERROR1}_{it} = & \beta_0 + \beta_1 \text{CC}_{it} + \beta_2 \text{SIZE}_{it} + \beta_3 \text{SURP}_{it} + \beta_4 \text{SURP}_{it-1} \\ & + \beta_5 \Delta \text{AGE1}_{it} + \beta_6 \text{FE}_{it\text{PRE}} + \varepsilon_{it} \end{aligned} \quad (1)$$

$$\begin{aligned} \text{ERROR2}_{it} = & \beta_0 + \beta_1 \text{CC}_{it} + \beta_2 \text{SIZE}_{it} + \beta_3 \text{SURP}_{it} + \beta_4 \text{SURP}_{it-1} \\ & + \beta_5 \Delta \text{AGE2}_{it} + \beta_6 \text{FE}_{it\text{PRE}} + \varepsilon_{it} \end{aligned} \quad (2)$$

$$\begin{aligned} \text{DISP1}_{it} = & \gamma_0 + \gamma_1 \text{CC}_{it} + \gamma_2 \text{SIZE}_{it} + \gamma_3 \text{SURP}_{it} + \gamma_4 \text{SURP}_{it-1} \\ & + \gamma_5 \Delta \text{AGE1}_{it} + \gamma_6 \text{DISP}_{it\text{PRE}} + \mu_{it} \end{aligned} \quad (3)$$

$$\begin{aligned} \text{DISP2}_{it} = & \gamma_0 + \gamma_1 \text{CC}_{it} + \gamma_2 \text{SIZE}_{it} + \gamma_3 \text{SURP}_{it} + \gamma_4 \text{SURP}_{it-1} \\ & + \gamma_5 \Delta \text{AGE2}_{it} + \gamma_6 \text{DISP}_{it\text{PRE}} + \mu_{it} \end{aligned} \quad (4)$$

CC is a dummy variable equal to 1 if the quarter is a conference call quarter and 0 otherwise. A significantly negative coefficient on β_1 in Equations (1) and (2) is consistent with conference calls reducing the *error* in analysts' earnings forecasts (P1). A significantly negative coefficient on γ_1 in Equations (3) and (4) is consistent with conference calls reducing the *dispersion* of analysts' earnings forecasts (P2).

Several features of our research design address potential alternative explanations. First, the use of the changes in forecast error and changes in forecast dispersion as the dependent variables reduces the potential effects of differences in firms' information environments, to the extent these differences affect the pre- and post-conference call measures equally. Second, as discussed above, the inclusion of control variables also reduces the possibility that differences in information environments affect our results. Third, the inclusion of non-conference call quarters controls for overall decreases in forecast error and dispersion as the earnings announcement date approaches. The use of this non-conference call control group also helps control for any general increase in the amount of information disclosed

¹⁵ The authors thank the referee for this suggestion.

in earnings announcements in recent years (Johnson et al. 2000). We also conduct a battery of sensitivity tests, which we report in Section IV.

Within-Firm, Across-Time Tests

We also perform within-firm, across-time comparisons of the four dependent measures for firms that begin using conference calls during our sample period. We compare the measures of error and dispersion during initial conference call quarters to the same non-conference call quarters one year earlier. If analysts immediately learn to use the information provided in conference calls, then we would expect the greatest decline in their forecast errors and in the dispersion of their forecasts to occur in the first conference call quarter. For completeness, we also examine the differences in the second and third conference call quarters. Note that it is possible that information released in earlier conference calls affects the pre-conference call measure (e.g., so that error decreased), which would reduce the potential for improvement for these subsequent conference call quarters.

For each firm with data available for both quarters, we compute the difference in the change in forecast error (and dispersion) between the first conference call quarter (or the 2nd or the 3rd conference call quarter) and the quarter one year earlier. We then test whether these differences are less than zero, i.e., whether error decreased more in conference call quarters. More specifically, we compute the differences in the four dependent variables as:

$$DERROR1_i = ERROR1_{it}^{CC} - ERROR1_{i,t-4}^{NCC} \quad (5)$$

$$DERROR2_i = ERROR2_{it}^{CC} - ERROR2_{i,t-4}^{NCC} \quad (6)$$

$$DDISP1_i = DISP1_{it}^{CC} - DISP1_{i,t-4}^{NCC} \quad (7)$$

$$DDISP2_i = DISP2_{it}^{CC} - DISP2_{i,t-4}^{NCC} \quad (8)$$

where i = firm, t = quarter, CC = conference call quarter, and NCC = non-conference call quarter.

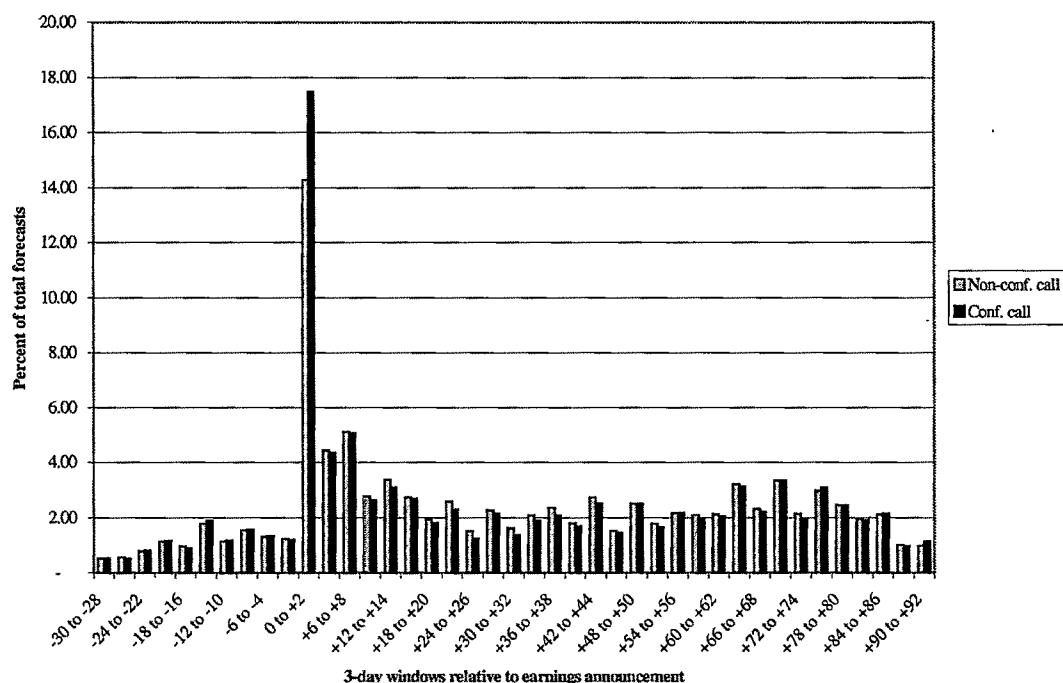
The within-firm, across-time analysis controls for firm-specific factors that are constant through time (rather than using non-conference call quarters as a control group). To the extent that firms do not change their other disclosure policies during that year, we expect ERROR1 (ERROR2) and DISP1 (DISP2) will be smaller (more negative) during the first conference call quarter relative to the same quarter of the prior year. Thus, we expect DERROR and DDISP will be negative, reflecting a greater decrease in forecast error and dispersion, respectively.

IV. EMPIRICAL RESULTS

Descriptive Statistics

Figure 2 plots the frequency of forecast revisions around earnings announcement dates for conference call vs. non-conference call quarters. Figure 2 shows a higher concentration of forecast revisions in the three-day window that corresponds to the day of and the two days after the earnings announcement in quarters when firms hold conference calls (17.5 percent) relative to those when firms do not hold conference calls (14.3 percent). The percentage of forecast revisions occurring in this three-day window (relative to those occurring during the rest of the quarter) is significantly higher for conference call quarters

FIGURE 2
Forecast Revisions—Conference Call vs. Non-Conference Call Quarters



than non-conference call quarters ($\chi^2(1) = 472.94$, $p = 0.001$).¹⁶ These data suggest that analysts revise their one-quarter-ahead forecasts closer to the earnings announcement date in quarters when firms host conference calls in conjunction with the earnings announcement, consistent with Frankel et al. (1999).

Table 3 provides descriptive statistics for the combined sample of conference call and non-conference call quarters. While the decreases in forecast error and dispersion appear small as a percent of share price, all are significantly different from zero. Further, as a percent of the pre-announcement-level of forecast error, the reductions in forecast error are approximately 12 percent and 26 percent for the first and second post-conference call windows, respectively, suggesting that forecast error decreases are also economically significant. The firms in the final sample are quite large with a mean (median) market value of equity of \$2,532 (\$632) million, falling in the 90th (70th) percentile of all firm-quarters reported on the quarterly Compustat files for the period 1995–98. Market values are also highly skewed, so our empirical tests use the log of market values. The negative value for $\Delta AGE1$ indicates that the average forecast is older in the pre-conference call measure than in the first post-conference call measure, while the positive value for $\Delta AGE2$ indicates that

¹⁶ We find similar results for one-day, two-day, and five-day windows. Most of the difference shown in Figure 2 occurs on day +1 (11.2 percent for conference call quarters vs. 8.9 percent for non-conference call quarters).

TABLE 3
Descriptive Statistics for Dependent, Independent, and Control Variables

	<i>n</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>Median</i>	<i>Minimum</i>	<i>Quartile1</i>	<i>Quartile3</i>	<i>Maximum</i>
Dependent variables								
ERROR1	12,032	-0.00083	0.00235	-0.00011	-0.02128	-0.00075	-0.00000	0.00419
ERROR2	10,286	-0.00181	0.00419	-0.00040	-0.03552	-0.00170	-0.00000	0.00500
DISP1	11,909	-0.00006	0.00121	-0.00000	-0.00690	-0.00030	0.00015	0.00681
DISP2	10,178	-0.00031	0.00144	-0.00006	-0.00897	-0.00060	0.00011	0.00615
Independent and control variables								
CC	12,555	0.3763	0.4845	0	1	1	0	0
SIZE	12,262	2,532.3	7,443.1	632.4	168,648.6	1,968.7	217.9	7.8
SURP _t	11,806	0.0132	0.0374	0.0047	2.2627	0.0114	0.0023	0.0000
SURP _{t-1}	11,551	0.0133	0.0486	0.0048	3.1936	0.0113	0.0023	0.0000
ΔAGE1	12,528	-9.1183	25.3660	-9.7	-80.0	-27.5	7.0	72.0
ΔAGE2	12,202	1.2204	25.2669	1.7	-82.0	-14.5	18.0	90.0
FE _{PRE}	12,032	0.00447	0.00901	0.00166	0.18286	0.00443	0.00068	0.00000
DISP _{PRE}	11,909	0.00137	0.00229	0.00069	0.05577	0.00154	0.00031	0.00000

$ERROR1 = (FE_{tPOST1} - FE_{tPRE})/p_{t-1}$
 $ERROR2 = (FE_{tPOST2} - FE_{tPRE})/p_{t-1}$
 $DISP1 = (DISP_{tPOST1} - DISP_{tPRE})/p_{t-1}$
 $DISP2 = (DISP_{tPOST2} - DISP_{tPRE})/p_{t-1}$

See Figure 1 for definitions of FE_{tPRE} , FE_{tPOST1} , FE_{tPOST2} , $DISP_{tPRE}$, $DISP_{tPOST1}$, and $DISP_{tPOST2}$;

CC = 1 if the firm hosts a conference call at quarter t-1 earnings announcement, 0 otherwise;

SIZE = log of market value of equity at the beginning of quarter t. For descriptive purposes, this table presents the actual market value (rather than the log values), but all subsequent tables use the log value;

SURP_t = $|EPS_t - EPS_{t-4}|/p_{t-4}$, where EPS_t = earnings per share including extraordinary items, p_{t-4} = ending price at quarter t-4;

ΔAGE1 = average age of forecasts of quarter t earnings 20 days after quarter t-1 earnings announcement (POST1) less the average age of forecasts of quarter t earnings at the beginning of the quarter (PRE). Both averages are based on the forecasts used to compute the dependent variables; and

ΔAGE2 = average age of forecasts of quarter t earnings announcement date (POST2) less the average age of forecasts of quarter t earnings at the beginning of the quarter (PRE). Both averages are based on the forecasts used to compute the dependent variables, ERROR2 and DISP2.

the average forecast is more recent in the pre-conference call measure than the second post-conference call measure. This finding is consistent with a concentration of forecast revisions following the earnings announcement.¹⁷

Table 4 provides the results of univariate tests comparing conference call quarters with non-conference call quarters. As predicted, both parametric t-tests and non-parametric Wilcoxon rank sum tests indicate that the decreases in error and dispersion of analysts' forecasts immediately after the earnings announcement (ERROR1 and DISP1) are significantly greater in quarters in which firms host conference calls in conjunction with the earnings announcement than in non-conference call quarters. The greater decrease in dispersion persists through the next earnings announcement (DISP2). Although the decrease in forecast error measured through the next earnings announcement (ERROR2) is not significant at conventional levels using a one-tailed parametric t-test ($p = 0.13$), the lack of significance is likely due to outliers, because the difference is highly significant using the non-parametric Wilcoxon rank sum test ($p = 0.0001$).¹⁸

Consistent with Tasker (1998) and Frankel et al. (1999), larger firms are more likely to host conference calls. The current and lagged earnings surprise variables ($SURP_t$ and $SURP_{t-1}$) are not significantly different between conference call and non-conference call quarters using a parametric t-test; however, the Wilcoxon rank sum test suggests that lagged surprise is higher in conference call quarters than non-conference call quarters, consistent with firms using conference calls to explain significant changes in earnings. The change in the average age of the forecasts ($\Delta AGE1$ and $\Delta AGE2$) is significantly greater in conference call quarters than non-conference call quarters, consistent with analysts revising more frequently in conference call quarters. Finally, the pre-conference call levels of forecast error and dispersion are smaller in conference call quarters than non-conference call quarters. To the extent that smaller pre-levels of error and dispersion result in smaller *changes* in error and dispersion, not controlling for the initial levels of error and dispersion weakens our ability to detect differences in each dependent variable for conference call vs. non-conference call firm-quarters.

Results of Cross-Sectional Tests

We present the results of the regressions to test P1, i.e., Equations (1) and (2) in Table 5, Panel A. We eliminate outliers using the procedures discussed in Belsley et al. (1980).¹⁹ As predicted, the coefficient on the conference call variable (CC) is significantly negative (p -values < 0.01) for both the shorter window (ERROR1) and the longer window (ERROR2). Thus, the decrease in forecast error directly after the earnings announcement is significantly greater in quarters in which firms hold conference calls in conjunction with the earnings announcement than in non-conference call quarters. These results further suggest that firms not hosting conference calls do not provide sufficient additional information

¹⁷ The average age of forecasts used to compute ERROR1(2) and DISP1(2) are 41 (43) days for the pre-conference call measure and 32 (44) days for the post-conference call measure. Thus, forecasts used to compute the first post-conference call measure (i.e., for ERROR1 and DISP1) tend to be more recent, reflecting the tendency for analysts to revise forecasts following the earnings announcement.

¹⁸ Large firms are more likely to use conference calls, i.e., the correlation between SIZE and CC is 0.21. Large firms also tend to have smaller changes in forecast error, i.e., the correlation between SIZE and ERROR1 and ERROR2 is -0.13 and -0.14 , respectively. This suggests that large conference call firms have less room for the conference call to further reduce error.

¹⁹ Specifically, we eliminate observations with studentized residuals with absolute value greater than 2, or with a hat matrix value greater than $2 \times p/n$ (where p = number of parameters and n = number of observations), or a DFFITS statistic greater than $2\sqrt{p/n}$. The main results are similar without this procedure, i.e., the coefficient on the CC variable is significant at less than the 2 percent level for all dependent variables.

TABLE 4
Univariate Tests Comparing Conference Call Quarters with Non-Conference Call Quarters

	Conference Call Quarters			Non-Conference Call Quarters			Prediction	Test of Mean Differences ^a		Non-Parametric Wilcoxon Rank Sum Test ^b	
	n	Mean	Median	n	Mean	Median		t-value	p-value	z-score	p-value
Dependent Variables ^c											
ERROR1	4,583	-0.00094	-0.00016	7,449	-0.00077	-0.00006	CC < NCC	3.8445	0.0001	6.4608	0.0001
ERROR2	4,046	-0.00186	-0.00044	6,240	-0.00177	-0.00037	CC < NCC	1.1448	0.1262	3.9841	0.0001
DISP1	4,530	-0.00010	0.00000	7,379	-0.00003	0.00000	CC < NCC	3.1177	0.0009	4.5930	0.0001
DISP2	4,003	-0.00035	-0.00012	6,175	-0.00028	0.00000	CC < NCC	2.4383	0.0074	4.5842	0.0001
Control Variables ^c											
SIZE	4,698	6.9354	6.8672	7,564	6.2727	6.1868	?	23.4074	0.0001	22.8535	0.0001
SURP _t	4,564	0.0138	0.0048	7,242	0.0129	0.0046	?	1.1211	0.2623	1.1640	0.2444
SURP _{t-1}	4,478	0.0133	0.0050	7,073	0.0134	0.0047	?	0.1549	0.8769	2.1366	0.0326
ΔAGE1	4,716	-12.0282	-13.3167	7,812	-7.3616	-7.6667	?	10.2838	0.0001	10.1783	0.0001
ΔAGE2	4,632	0.4265	0.5000	7,570	1.7062	2.6667	?	2.7429	0.0031	3.7264	0.0001
FE _{PRE}	4,583	0.00399	0.00150	7,449	0.00477	0.00177	?	4.8979	0.0001	5.3864	0.0001
DISP _{PRE}	4,530	0.00130	0.00068	7,379	0.00141	0.00069	?	2.7809	0.0054	1.5071	0.1318

^a Two-sample t-test (with adjustment for unequal variances where appropriate). p-values reported are one-tailed for dependent variables and two-tailed for control variables.

^b p-values reported are one-tailed for dependent variables and two-tailed for control variables.

^c See Table 3 for variable definitions.

^a Two-sample t-test (with adjustment for unequal variances where appropriate). p-values reported are one-tailed for dependent variables and two-tailed for control variables.

^b p-values reported are one-tailed for dependent variables and two-tailed for control variables.

^c See Table 3 for variable definitions.

TABLE 5

Cross-Sectional Regressions: Are Conference Calls Associated with Changes in Forecast Error and Changes in Forecast Dispersion?

Panel A: Changes in Forecast Error

P1: The decrease in forecast error will be greater in conference call quarters than in non-conference call quarters ($\beta_1 < 0$).

$$\text{Model: } \text{ERROR1}(2)_{it} = \beta_0 + \beta_1 \text{CC}_{it} + \beta_2 \text{SIZE}_{it} + \beta_3 \text{SURP}_{it} + \beta_4 \text{SURP}_{it-1} + \beta_5 \Delta \text{AGE1}(2)_{it} + \beta_6 \text{FE}_{it\text{PRE}} + \varepsilon_{it}$$

Variable ^a	ERROR1		ERROR2	
	Coefficient	p-value ^b	Coefficient	p-value ^b
Intercept	0.0000770	0.0946	0.0006930	0.0001
CC	-0.0001330	0.0001	-0.0002270	0.0001
SIZE	0.0000078	0.2288	-0.0000426	0.0001
SURP _t	-0.0017750	0.0095	0.0024710	0.0103
SURP _{t-1}	-0.0029040	0.0001	0.0016270	0.0529
$\Delta \text{AGE1}(2)$	0.0000058	0.0001	0.0000051	0.0001
FE _{PRE}	-0.1483930	0.0001	-0.5496530	0.0001
Adjusted R ² = 0.3126			Adjusted R ² = 0.7070	
n = 10,079 firm-quarters			n = 8,642 firm-quarters	

Panel B: Changes in Forecast Dispersion

P2: The decrease in forecast dispersion will be greater in conference call quarters than in non-conference call quarters ($\gamma_1 < 0$).

$$\text{Model: } \text{DISP1}(2)_{it} = \gamma_0 + \gamma_1 \text{CC}_{it} + \gamma_2 \text{SIZE}_{it} + \gamma_3 \text{SURP}_{it} + \gamma_4 \text{SURP}_{it-1} + \gamma_5 \Delta \text{AGE1}(2)_{it} + \gamma_6 \text{DISP}_{it\text{PRE}} + \mu_{it}$$

Variable ^a	DISP1		DISP2	
	Coefficient	p-value ^b	Coefficient	p-value ^b
Intercept	0.0002420	0.0001	0.0004460	0.0001
CC	-0.0000394	0.0022	-0.0000782	0.0001
SIZE	-0.0000094	0.0241	-0.0000217	0.0001
SURP _t	0.0024040	0.0001	0.0031020	0.0001
SURP _{t-1}	0.0017990	0.0001	0.0016530	0.0002
$\Delta \text{AGE1}(2)$	0.0000019	0.0001	0.0000002	0.5788
DISP _{PRE}	-0.2474290	0.0001	-0.5844560	0.0001
Adjusted R ² = 0.1825			Adjusted R ² = 0.5218	
n = 9,813 firm-quarters			n = 8,524 firm-quarters	

^a See Table 3 for variable definitions.

^b The p-values reported are two-tailed except for the CC variable, which is one-tailed.

later in the quarter to “catch up” with firms that host conference calls. Using the shorter window (ERROR1), the decrease in error appears to be about 0.013 percent (of price) greater in conference call quarters than non-conference call quarters. While it is difficult to assess the informational advantage that this result represents, a comparison of the difference to the empirical distribution of ERROR1 provides a sense of the economic significance of the difference. The interquartile range for ERROR1 (in the sample used for the regression analysis) is 0.062 percent. Thus, the difference is approximately 21 percent of the interquartile range. A similar assessment for the longer window indicates that the decrease in error for conference call firms is approximately 0.022 percent (of price) or 17 percent of the interquartile range. Overall, our results suggest that firms disclose more information during conference call quarters than in non-conference call quarters and that this information difference persists through the next quarter’s earnings announcement.

Table 5, Panel B reports the results of the regressions to test P2 (Equations [3] and [4]). As predicted, the coefficient on the conference call variable is significantly negative (p -values < 0.01) for both the shorter window (DISP1) and the longer window (DISP2). These results are consistent with our prediction that the decrease in forecast dispersion in the period after the prior quarter’s earnings announcement is significantly greater in quarters in which firms host conference calls in conjunction with the earnings announcement than in non-conference call quarters. For DISP1, the difference of 0.004 percent (of price) is approximately 11 percent of the interquartile range; for DISP2, the magnitude of the effect is approximately 0.007 percent of price and 13 percent of the interquartile range. Overall, our results suggest that dispersion in analysts’ forecasts decreases more during conference call quarters than in non-conference call quarters, and that this decrease persists through the next quarter’s earnings announcement.²⁰

Sensitivity Analyses for Cross-Sectional Regressions

In addition to the previously discussed features of our research design that reduce the possibility that our results primarily reflect differences across firms’ information environments (e.g., our use of changes in forecast accuracy vs. levels, and the incorporation of firm size and pre-announcement levels of forecast error and dispersion), we perform one final procedure. We add to the regression models a dummy variable (CCFIRM) that equals 1 for all firm-quarters (both conference call and non-conference call) for firms that we identify as regular conference call users:

$$\begin{aligned} \text{ERROR1}_{it} = & \beta_0 + \beta_1 \text{CC}_{it} + \beta_2 \text{CCFIRM}_{it} + \beta_3 \text{SIZE}_{it} + \beta_4 \text{SURP}_{it} + \beta_5 \text{SURP}_{it-1} \\ & + \beta_6 \Delta \text{AGE1}_{it} + \beta_7 \text{FE}_{it\text{PRE}} + \varepsilon_{it} \end{aligned}$$

The coefficient on CCFIRM is the change in error in the non-conference call quarters of firms that eventually host at least one conference call, relative to the change in error for non-conference call quarters of firms that never host conference calls. More importantly, the interpretation of the coefficient on CC changes—it now represents the incremental effect on analysts’ forecasts of hosting a conference call in a particular quarter for firms that regularly host conference calls. In this specification, the coefficient on CC is significant for

²⁰ Control variables are discussed in Section III. While we have no prediction for the sign of the coefficients, all except SIZE in the ERROR1 regression and ΔAGE2 in the DISP2 regression are significant. In particular, FE_{PRE} and DISP_{PRE} are highly significant in all regressions, consistent with the notion that larger pre-announcement levels of forecast error and dispersion allow for a greater reduction in forecast error and dispersion during the quarter.

all dependent variables at the 1 percent level or better. The coefficient on CCFIRM is generally not significant (except in the DISP2 regression), suggesting that non-conference call quarters for firms that eventually host conference calls are not significantly different from non-conference call quarters for firms that never host conference calls. This analysis suggests that firm-specific differences in information environments between firms that regularly host conference calls and those that do not are not driving our main results.

We also perform a number of other sensitivity checks on the cross-sectional regressions. Relying on t-statistics based on White's consistent covariance matrix yields results that are virtually identical to those reported in the tables. We also repeat the regressions after separately including (1) calendar quarter dummy variables to control for potential time-period specific effects, (2) dummy variables to control for potential fiscal-quarter effects, (3) two-digit SIC code dummy variables to control for potential industry effects, and (4) a time-trend variable to control for potential increases in disclosures in recent years. In all cases, the coefficient on the CC variables remains significant at better than the 5 percent level.

Next, we test the sensitivity of our results to our selection of conference call quarters. The last three sample selection procedures (summarized in Table 2) were designed to increase the power of our tests. If we eliminate these last three selection criteria, then the coefficient on the CC variable remains significant at the 5 percent level for all dependent variables. In addition, if we redefine regular users of conference calls as those firms that host conference calls in at least 75 percent of the quarters (after their first conference call quarter), then the coefficient on the CC variable again remains significant at the 5 percent level for all dependent variables.²¹

Finally, although changes in forecast error and dispersion are less likely to be correlated across time for a given firm than levels of forecast error and dispersion, we conduct two tests to address potential dependence in our pooled data. In the first test, we re-estimate our regressions after averaging each dependent and independent variable by firm, so there is only one observation per firm in this analysis. In the second test, we randomly select one observation per firm and re-run the regressions on this subset of observations. In both tests, inferences are not changed, suggesting that multiple observations for the same firm are not significantly biasing our results.

Results of Within-Firm Tests

Table 6 presents the results of the within-firm tests. The first set of four columns presents statistics on the (univariate) differences in the four dependent (and control) variables between the first conference call quarter and the same quarter one year prior. We report the mean, median, and percent of observations less than zero. The reported p-values are one-sample t-tests and non-parametric Wilcoxon signed rank tests, respectively. As predicted, the decrease in forecast error (DERROR1 and DERROR2) is greater in the first conference call quarter than in the same quarter one year prior (all p-values < 0.02). Of the firms with a non-zero difference between the first-conference call quarter and the same quarter one year prior, between 53 percent and 55 percent show a greater decrease in error during the conference call quarter.

The next two sets of four columns present similar results for the second and third conference call periods compared to those for the same quarters one year prior. The results for these quarters are somewhat weaker—as might be expected if the first conference call

²¹ As another sensitivity check, we also eliminated all non-conference call observations for any firm that eventually hosts a conference call, per the First Call database (i.e., both sporadic and regular conference call users). We find similar results.

TABLE 6
Within-Firm Analyses: Are Conference Calls Associated with Changes in Forecast Error and Changes in Dispersion?

	Difference between First Conference Call Quarter and Same Quarter One Year Prior				Difference between Second Conference Call Quarter and Same Quarter One Year Prior				Difference between Third Conference Call Quarter and Same Quarter One Year Prior			
	<i>n</i>	<i>Mean^a</i>	<i>Median^b</i>	% < 0 ^c	<i>n</i>	<i>Mean^a</i>	<i>Median^b</i>	% < 0 ^c	<i>n</i>	<i>Mean^a</i>	<i>Median^b</i>	% < 0 ^c
Dependent variables ^d												
DEROR1	370	-0.00041 p = 0.0008	0.00000 p = 0.0075	53.43%	446	-0.00028 p = 0.0138	-0.00003 p = 0.0366	55.61%	286	-0.00025 p = 0.0673	0.00000 p = 0.1722	50.75%
DEROR2	326	-0.00054 p = 0.0075	-0.00008 p = 0.0139	54.72%	394	-0.00059 p = 0.0051	-0.00001 p = 0.2815	51.67%	255	-0.00069 p = 0.0145	-0.00003 p = 0.1191	51.60%
DDISP1	364	-0.00002 p = 0.4162	0.00000 p = 0.7148	48.55%	440	-0.00001 p = 0.4383	0.00000 p = 0.2714	52.07%	276	-0.00008 p = 0.2307	0.00000 p = 0.1815	51.72%
DDISP2	319	0.00007 p = 0.7462	0.00001 p = 0.8104	47.74%	388	-0.00002 p = 0.4224	-0.00002 p = 0.2182	53.05%	251	-0.00017 p = 0.0782	-0.00005 p = 0.0891	53.41%

^a p-values reported are from one sample t tests of the means and are one-tailed for the dependent variables and two-tailed for the control variables.

^b p-values reported are from a Wilcoxon sign-rank test and are one-tailed for the dependent variables and two-tailed for the control variables.

^c Represents the proportion of negative values.

^d The dependent variables are the difference between each variable (ERROR1, ERROR2, DISP1, DISP2) measured during the first, second, or third conference call quarter and the same quarter one year prior. See Table 3 for definitions of ERROR1, ERROR2, DISP1, and DISP2.

decreased the forecast error in the pre-conference call period for the second and third conference call quarters. The means and medians are all negative (indicating a larger decrease in error during the conference call quarter compared to that from one year prior); however, the differences tend to become less significant with each quarter. Overall, the results reported here support the previous cross-sectional results and suggest that conference calls increase the information available to analysts that assists them in accurately forecasting a firm's earnings for the next quarter.

However, inconsistent with the results of our cross-sectional tests, the results of the within-firm tests for forecast dispersion do not support P2. The decrease in forecast dispersion around the prior quarter's earnings announcement is not significantly greater during the first conference call quarter compared to that for the same quarter one year earlier. We also find no significant differences for the second and third conference call quarters. The insignificant results of the within-firm forecast dispersion tests may call into question our cross-sectional test results on dispersion. However, the smaller sample sizes in the within-firm tests reduce our ability to detect differences between conference call and non-conference call quarters, should they exist.

Sensitivity Analyses on Within-Firm Tests

A potential problem for the within-firm analysis is that we cannot ascertain whether a firm's first appearance on the database is truly the firm's first conference call. Firms whose first appearance in our sample coincides with the start of the database may have held conference calls in earlier quarters. As a result, the comparison quarter from the prior year, which we deem a non-conference call quarter, might in fact have been an unobserved conference call quarter, thereby weakening our tests. Since the potential for this problem is greatest for firms whose first conference call quarter is at the start of the database, we perform our within-firm tests separately for this group of firms (deemed the "early-adopters") and on the subset of firms whose first conference call quarter occurs after the start of the database (deemed the "late-adopters"). Our results (not reported in the tables) are strong for the early-adopters and weak for the late-adopters. The fact that the results are not weaker (and in fact are stronger) for the early adopters suggests that any potential misclassification caused by considering the first conference call quarter on the database as the first conference call the firm held does not materially affect our inferences.

Another potential problem for the within-firm research design is that any increase in the tendency to disclose forward-looking information in recent years could explain the difference in the changes in error and dispersion between the first conference call quarter and the same quarter one year prior. To address this possibility, we perform similar within-firm tests on a sample of non-conference call firms matched to our conference call adopters on quarter, two-digit SIC code, and market value. The results of these tests (not reported in the tables) indicate that non-conference call firms do not experience similar decreases in error over the same time period (all p-values > 10 percent). This fact in turn suggests that a general increase in the tendency to disclose forward-looking information in recent years does not drive the results for our conference call adopters.

Finally, our within-firm design does not control for potential differences across *quarters*. Of the control variables included in our cross-sectional tests, only the SURP variables attempt to control for quarter-specific factors that could affect our results. The remaining control variables address cross-sectional differences that should be controlled for by our within-firm design. Thus, we run regressions on our within-firm sample with DERROR and DDISP on the left-hand side and $DSURP_t$ and $DSURP_{t-1}$ on the right-hand side. The intercepts from these regressions represent the mean value of DERROR and DDISP after

controlling for earnings surprises in the current and prior quarters. Similar to results from the univariate tests reported in Table 6, the intercepts in the DERROR regressions are significant, consistent with conference calls reducing the error in analysts' forecasts. Also consistent with the univariate results, the intercepts in the DDISP regressions were not significant.

V. DO CONFERENCE CALLS BENEFIT ALL ANALYSTS EQUALLY?

Both our cross-sectional tests and our within-firm tests support the conclusion that conference calls provide information useful to analysts in predicting future earnings. However, it is possible that the information managers provide during conference calls does not benefit all analysts equally. In other words, a subset of analysts may be driving the decrease in error. We investigate this issue by comparing the changes in individual analysts' forecasts error across different types of analysts.

First, we partition analysts into two groups based on forecasting ability. Stickel (1992) and Sinha et al. (1997) find a positive relation between past forecast accuracy and current forecast accuracy. Brown (2001) finds that a simple model including only past forecast accuracy explains current forecast accuracy as well as a relatively complex model incorporating economic determinants of forecast accuracy identified in the prior literature. Thus, we use a relative ranking of analysts' prior forecast accuracy as a proxy for ability. We begin by computing mean-adjusted forecast errors by deducting the average forecast error for the firm quarter from each individual analyst's forecast error. We then compute the average mean-adjusted forecast error for each analyst over the prior two years, and rank the analysts based on this average. We compute percentage rank as the rank value divided by the number of analysts ranked.²² We consider analysts with rankings greater than 0.5 to be "high-ability" analysts and analysts with rankings less than or equal to 0.5 to be "low-ability" analysts. For each firm-quarter, we compute the average ERROR1 and ERROR2 variable for the high-ability (ERROR1^{HI} and ERROR2^{HI}) and low-ability (ERROR1^{LO} and ERROR2^{LO}) groups. We also compute the difference between the two groups (ERROR1^{HI-LO} and ERROR2^{HI-LO}) for firms with at least one analyst in each group. If conference calls level the playing field across analysts, then the improvement in accuracy for the low-ability group will exceed the improvement for the high-ability group, and ERROR1(2)^{HI-LO} will be larger (less negative) in conference call quarters. On the other hand, if the high-ability group makes better use of the information provided in the conference call, then ERROR1(2)^{HI-LO} will be smaller (more negative) in conference call quarters. Given the two opposing predictions, we conduct two-tailed tests.

Table 7, Panel A presents univariate comparisons of the decrease in forecast error between conference call quarters and non-conference call quarters. The results are analogous to those in Table 4 on ERROR1 and ERROR2—only in this table we have decomposed the ERROR1 and ERROR2 variables into two groups based on analyst ability. Conference calls appear to have a short-term benefit for both high- and low-ability analysts, i.e., ERROR1^{HI} and ERROR1^{LO} are both smaller (more negative) in conference call quarters than non-conference call quarters (p-values < 0.05). However, the difference between the high- and low-ability groups (ERROR1^{HI-LO}) is larger (less negative) in conference call quarters than non-conference call quarters (p-values < 0.05 for both t-test and non-parametric Wilcoxon rank sum test), consistent with conference calls providing greater

²² These rankings are based on all analysts contributing forecasts for a given firm on Zack's individual estimate history, whereas the analyst forecasts that comprise our computation of ERROR1(2) must meet other data requirements (e.g., must have forecasts in both the pre- and post-conference call periods). Thus, the partitioning variable will not necessarily result in an even split of the forecasts comprising ERROR1(2) and we will lose some firm-quarter observations if we can consider no analysts to be high or low within that firm-quarter.

TABLE 7
Do Conference Calls Benefit All Analysts Equally?
Univariate Tests of Forecast Error by Analyst Type in Conference Call vs.
Non-Conference Call Quarters

	Conference Call Quarters			Non-Conference Call Quarters			Test of Mean Differences ^a		Non-Parametric Wilcoxon Rank Sum Test ^b	
	n	Mean	Median	n	Mean	Median	t-value	p-value	z-score	p-value
Panel A: Partition Based on Analysts' Prior Forecast Error ^c										
ERROR1:										
ERROR1 ^{HI}	3,553	-0.00097	-0.00010	4,977	-0.00083	-0.00000	2.113	0.0347	6.367	0.0001
ERROR1 ^{LO}	3,535	-0.00093	-0.00000	4,949	-0.00064	-0.00000	4.998	0.0001	8.380	0.0001
ERROR1 ^{HI-LO}	3,488	-0.00004	-0.00000	4,881	-0.00018	-0.00000	2.173	0.0298	2.160	0.0308
ERROR2:										
ERROR2 ^{HI}	3,151	-0.00186	-0.00042	4,220	-0.00182	-0.00030	0.390	0.6967	4.291	0.0001
ERROR2 ^{LO}	3,130	-0.00192	-0.00043	4,187	-0.00167	-0.00026	2.316	0.0206	5.400	0.0001
ERROR2 ^{HI-LO}	3,007	0.00006	-0.00000	4,023	-0.00012	-0.00000	2.068	0.0387	2.322	0.0202
Panel B: Partition Based on Analysts' Brokerage House Affiliation ^d										
ERROR1:										
ERROR1 ^{HI}	3,443	-0.00100	-0.00011	4,549	-0.00084	0.00000	2.309	0.0210	5.205	0.0001
ERROR1 ^{LO}	3,457	-0.00097	-0.00006	4,543	-0.00072	0.00000	3.725	0.0002	7.663	0.0001
ERROR1 ^{HI-LO}	3,412	-0.00003	0.00000	4,488	-0.00013	0.00000	1.304	0.1921	1.807	0.0707
ERROR2:										
ERROR2 ^{HI}	3,038	-0.00187	-0.00042	3,869	-0.00183	-0.00033	0.316	0.7518	2.835	0.0046
ERROR2 ^{LO}	3,043	-0.00194	-0.00043	3,957	-0.00171	-0.00028	2.060	0.0395	4.532	0.0001
ERROR2 ^{HI-LO}	2,907	0.00007	0.00000	3,753	-0.00007	0.00000	1.463	0.1436	1.664	0.0961

(Continued on next page)

TABLE 7 (Continued)

^a Two-sample t-test (with adjustment for unequal variances where appropriate). p-values reported are one-tailed for $ERROR1(2)^{HI}$ and $ERROR1(2)^{LO}$ and two-tailed for the difference variable, $ERROR1(2)^{HI-LO}$.
^b p-values reported are one-tailed for $ERROR1(2)^{HI}$ and $ERROR1(2)^{LO}$ and two-tailed for the difference variable, $ERROR1(2)^{HI-LO}$.
^c $ERROR1(2)^{HI}$ and $ERROR1(2)^{LO}$ represent the average of $ERROR1(2)$ for the high-ability and low-ability analysts as classified based on prior forecast error. $ERROR1(2)^{HI-LO}$ represents the difference between $ERROR1(2)^{HI}$ and $ERROR1(2)^{LO}$. The partition between high-ability and low-ability analysts is based on a relative ranking of prior average (mean-adjusted) forecast errors (computed over the prior two years) for all analysts with forecasts for a given firm on Zack's individual analyst estimate history. High-ability analysts have relative rankings greater than 50 percent (with at least two quarters of prior forecasts); low-ability analysts have relative rankings equal to or less than 50 percent.
^d Same as note c above except the partition is based on whether the analyst is associated with a brokerage house listed as one of the top 20 research teams between 1995 and 1998, by <i>Institutional Investor</i> magazine. Footnote 23 in the text lists these top brokerage houses.

benefit to low-ability analysts. For the longer window, the difference between conference call quarters and non-conference call quarters is not significant for the high-ability group ($ERROR2^{HI}$) using a t-test ($p = 0.379$), but is significant using the Wilcoxon test ($z = 4.30$; $p = 0.0001$). The effect of conference calls for the low-ability group ($ERROR2^{LO}$), however, is stronger (t-stat = 2.32, $p = 0.021$; Wilcoxon z-score = 5.40, $p = 0.0001$). Once again, the difference between the high and low groups ($ERROR2^{HI-LO}$) is larger in conference call quarters than non-conference call quarters (p -values < 0.05 for both t-test and Wilcoxon rank sum test). In summary, while both high- and low-ability analysts benefit from conference calls, low-ability analysts appear to benefit more.

Table 8, Panel A presents the results of cross-sectional regressions of the difference variable, $ERROR1(2)^{HI-LO}$, on a conference call dummy variable and on the control variables used in the previous regressions. If conference calls benefit the low-ability group more than the high-ability group, then $ERROR1(2)^{HI-LO}$ should be larger in conference call quarters and the coefficient on CC should be positive. For these regressions, we computed the $\Delta AGE1(2)$ variable and the FE_{PRE} variable at the individual-analyst level and aggregate them between the high- and low-ability groups. Similar to the dependent variable, the regression includes the difference in these variables between the high- and low-ability groups. The coefficient on CC is significantly positive for both the shorter window ($ERROR1^{HI-LO}$, $p = 0.014$, two-tailed) and the longer window ($ERROR2^{HI-LO}$, $p = 0.031$, two-tailed). Similar to our univariate results, these findings suggest that low-ability analysts benefit more (in terms of improvement in forecast error) from conference calls than do high-ability analysts.

We also perform similar analyses partitioning analysts based on their brokerage-firm affiliations. Using *Institutional Investor's* ranking of research teams from various brokerage firms, we identify analysts associated with firms that are listed in the top 20 at least once during the period 1995–98.²³ We consider analysts from these firms to be of higher quality and/or to have greater access to managers. Table 7, Panel B presents the results of differences between high- and low-quality analysts. Similar to the results presented in Panel A, conference calls appear to benefit both high- and low-quality analysts—particularly in the shorter window. However, while the effect of conference calls appears to be stronger for the low-quality analysts, the difference between the high- and low-quality groups is not statistically different between conference call and non-conference call quarters using means (p -values > 0.10) and only marginally significant based on the Wilcoxon test ($p = 0.07$ for $ERROR1^{HI-LO}$ and $p = 0.10$ for $ERROR2^{HI-LO}$). Parallel to the analysis based on analyst ability, Panel B of Table 8 presents the results of cross-sectional regressions of the difference variable, $ERROR1(2)^{HI-LO}$, on a conference call dummy variable and control variables. The coefficient on the CC variable is not significant at conventional levels for the shorter window and approaches significance in the longer windows ($p = 0.104$).²⁴ Overall, the results on analyst type based on brokerage firm affiliation provide, at best, weak evidence that conference calls provide greater benefits to lower ranked analysts.²⁵

²³ Firms listed in the top 20 at least once during the 1995–98 period include: Merrill Lynch; Morgan Stanley Dean Witter; Goldman Sachs; Donaldson, Lufkin, and Jenrette; Saloman Smith Barney; Paine Webber; CS First Boston; Bear Stearns; Prudential Securities; Sanford C. Bernstein; Lehman Brothers; Schroder & Co.; Cowen & Co.; JP Morgan; BT Alex Brown; Deutsche Bank Securities; CIBC Oppenheimer; Nations Bank Montgomery Securities; Warburg Dillon Reed; NatWest Securities; and Gerard Klauer Mattison.

²⁴ We performed these tests using a within-firm design (similar to the tests presented in Table 6); however the sample sizes are small (less than 200 for the difference variable $ERROR1(2)^{HI-LO}$). The results do not indicate that the high- and low-ability (quality) groups are significantly different before and after the initial conference call.

²⁵ Except for SIZE, all of the control variables in Table 8 were significant.

TABLE 8
Do Conference Calls Benefit All Analysts Equally?
Cross-Sectional Regression of Difference in Forecast Error between
High-Ability and Low-Ability Analysts

$$\text{Model: } \text{ERROR1}(2)_{it}^{\text{HI-LO}} = \beta_0 + \beta_1 \text{CC}_{it} + \beta_2 \text{SIZE}_{it} + \beta_3 \text{SURP}_{it} + \beta_4 \text{SURP}_{it-1} \\ + \beta_5 \Delta \text{AGE1}(2)_{it}^{\text{HI-LO}} + \beta_6 \text{FE}_{it\text{PRE}}^{\text{HI-LO}} + \varepsilon_{it}$$

Panel A: Partition Based on Prior Forecast Error^a

<i>Variable^b</i>	<i>ERROR1^{HI-LO}</i>		<i>ERROR2^{HI-LO}</i>	
	<i>Coefficient</i>	<i>p-value^c</i>	<i>Coefficient</i>	<i>p-value^c</i>
Intercept	-0.000068	0.2777	-0.000035	0.5879
CC	0.000065	0.0142	0.000058	0.0310
SIZE	0.000004	0.6267	-0.000006	0.5185
SURP _{<i>t</i>}	-0.002338	0.0070	-0.001864	0.0461
SURP _{<i>t-1</i>}	-0.004068	0.0001	-0.003560	0.0001
$\Delta \text{AGE1}(2)_{it}^{\text{HI-LO}}$	0.004570	0.0001	0.000002	0.0001
$\text{FE}_{it\text{PRE}}^{\text{HI-LO}}$	-0.415812	0.0001	-0.860780	0.0001
Adjusted R ² = 0.2987			Adjusted R ² = 0.6504	
n = 7,483 firm-quarters			n = 6,314 firm-quarters	

Panel B: Partition Based on Brokerage House Association^a

<i>Variable</i>	<i>ERROR1^{HI-LO}</i>		<i>ERROR2^{HI-LO}</i>	
	<i>Coefficient</i>	<i>p-value^c</i>	<i>Coefficient</i>	<i>p-value^c</i>
Intercept	-0.000099	0.1758	-0.000088	0.2151
CC	0.000026	0.3696	0.000045	0.1045
SIZE	0.000010	0.2970	0.000002	0.8299
SURP _{<i>t</i>}	-0.004912	0.0001	-0.000991	0.3918
SURP _{<i>t-1</i>}	-0.002132	0.0281	-0.002066	0.0260
$\Delta \text{AGE1}(2)_{it}^{\text{HI-LO}}$	0.000004	0.0001	0.000003	0.0001
$\text{FE}_{it\text{PRE}}^{\text{HI-LO}}$	-0.445257	0.0001	-0.880104	0.0001
Adjusted R ² = 0.3081			Adjusted R ² = 0.6654	
n = 6,735 firm-quarters			n = 5,727 firm-quarters	

^a Notes c and d in Table 7 describe the partition.

^b $\Delta \text{AGE1}(2)_{it}^{\text{HI-LO}}$ is the difference between the average $\Delta \text{AGE1}(2)$ for the high-ability analysts and the low-ability analysts. See Table 3 for definition of $\Delta \text{AGE1}(2)$. Notes to Figure 1, Table 3, and Table 7 contain definitions of other variables.

^c The p-values reported are two-tailed.

VI. SUMMARY AND CONCLUSIONS

Substantial growth in the use of conference calls in recent years, combined with the SEC's concern about the potential for selective disclosure in conference calls, motivated us to examine the change in the error and dispersion of analysts' forecasts following earnings-related conference calls.

Our evidence suggests that the regular use of earnings-related conference calls increases the amount of public information available to analysts, improving their ability to accurately forecast earnings for the next quarter. In cross-sectional tests, we find that the decrease in analysts' forecast errors directly after the prior quarter's earnings announcement is greater when firms host conference calls than when they do not. This greater decrease in forecast error persists until the subsequent earnings announcement. We also conduct within-firm tests, comparing the change in forecast error for the same firm before and after the firm starts hosting conference calls, and again find a greater decrease in forecast error when the firm first hosts conference calls. Our results suggest that managers disclose more information during conference call quarters than during non-conference call quarters, and that this information difference persists through the next quarter's earnings announcement.

We also find some evidence that conference calls tend to reduce the dispersion in analysts' forecasts, although these results are less conclusive. In our cross-sectional tests, we find that the decrease in the dispersion of analysts' forecasts directly after the prior quarter's earnings announcement is relatively greater in quarters when firms hold conference calls than in quarters when they do not. However, our within-firm comparisons do not provide similar support. The dispersion in analysts' forecasts does not decrease significantly after the firm begins to host conference calls (but our within-firm tests are necessarily based on much smaller samples, reducing the power of our tests to detect differences).

Overall, our results suggest that conference calls could present a selective disclosure problem if the public is not privy to these calls. During our sample period (prior to the passage of Reg FD), relatively few firms allowed individuals to participate in these calls—a 1998 survey by NIRI indicated only 29 percent of firms surveyed invited individuals to participate in their conference calls and that only 14 percent invited the media (NIRI 1998c). While the passage of Reg FD has likely increased the proportion of firms that allow open access to their calls, there is still considerable controversy over the effect of Reg FD on the information flow to the market (see for example, Opdyke and Schroeder 2001). Given the public had, at best, restricted access to conference calls during our sample period, our evidence is consistent with these calls increasing the information provided to financial analysts. Our evidence further suggests that an information gap existed between the analyst community and the remainder of the investment community, which is consistent with the SEC's concerns.

We also consider whether conference calls “level the playing field” across analysts by examining whether conference calls provide more benefits to analysts with relatively weaker forecasting ability and/or to those with less access to management. Using prior forecast accuracy as a proxy for analysts' forecasting ability, our results indicate that both low- and high-ability analysts benefit from conference calls. However, analysts with weaker prior forecasting ability benefit more from these calls, suggesting that conference calls do act to level the playing field between these two groups. Our results on the differential effects of conference calls on the forecasts of analysts with less access to management are inconclusive.

Our study is subject to several limitations. First, our sample only examines conference calls held regularly in conjunction with quarterly earnings announcements. Second, the sample consists of relatively large firms. Thus, the results may not generalize to other types

of conference calls or to the broader population of firms. Third, we cannot determine which of our sample conference calls were restricted from public access, although given the time period covered in the sample, most calls were likely restricted. Finally, we cannot rule out the possibility that these firms disclose the same information introduced during the analyst conference call simultaneously to the entire investment community, although it seems unlikely that a press release could fully capture the richness of a conference call.

APPENDIX

Implications of BKLS Framework for Forecast Error and Dispersion

In Barron et al. (1998) (hereafter BKLS), analysts use both public and private information in forecasting firms' earnings. They denote the precision of public information as h and the precision of private information as s . Their framework focuses on two empirically observable properties of analysts' forecasts—forecast dispersion (D) and the squared error in the mean forecast (SE). In the simple case where the precision of private information is the same across analysts ($s_i = s$ for all i), forecast dispersion and squared error are expressed in terms of h and s as follows:

$$D = \frac{s}{(h + s)^2} \quad SE = \frac{h + s/N}{(h + s)^2}$$

Thus, one can express the effect of the precision of public information (h) on dispersion (D) and squared error (SE) of analysts' forecasts as follows:

$$\frac{\partial D}{\partial h} = \frac{-2s}{(h + s)^3} < 0 \quad \frac{\partial SE}{\partial h} = \frac{s - (2s/N) - h}{(h + s)^3} < 0 \quad \text{if } h > s \left(\frac{N - 2}{N} \right)$$

The above analysis implies that the relation between forecast dispersion and the precision of public information is negative, i.e., that forecast dispersion decreases as h increases. The relation between squared error and the precision of public information is not as simple. The relation is negative (i.e., that forecast error decreases as h increases) if public information is at least as precise as private information (or if h is at least greater than $s \times [(N - 2)/N]$). Intuitively, increased precision in public information increases the extent to which analysts rely on public information relative to private information. Because private information is (by definition) idiosyncratic, the error in private signals tends to cancel out when averaging across a number of analysts (i.e., there is a "diversification effect"). Thus, if public information is imprecise relative to private information, then the benefit of increased precision in the public signal is offset by a loss in the benefits received from diversification because analysts rely less on their private signals, and the net effect is an increase in forecast error. However, when the precision of public information is high relative to the precision of private information, the loss of benefit from diversification is small and the increased precision in the public information should reduce the error in analysts' forecasts.

We expect the effect of increased precision of public information to outweigh the effect of reduced diversification because we argue that public information is, on average, more precise than private information (i.e., that $h > s$). Public information provided by the firm such as annual reports, quarterly filings, and earnings announcements, are important sources of information. In addition, given that in the BKLS framework private information is uncorrelated across analysts, the BKLS framework would consider as public any common

information managers provide to analysts in private conversations (and it is unlikely that the information managers provide would be completely "uncorrelated" across analysts). As such, much of the information conveyed in private conversations between managers and analysts is most likely "public" information in the BKLS framework.

Our analysis holds the precision of private information, s , constant. The effect of conference calls on private information is ambiguous. Conference calls may increase the precision of private information if analysts process the information management releases during a conference call into new private information (e.g., Kim and Verrechia 1994). Alternatively, conference calls could reduce the precision of private information if they reduce analysts' incentives to gather private information (because public information mitigates the gains from private information). Thus, we hypothesize an effect only for public information, and base our predictions on the assumption that any changes in private information are immaterial second-order effects.

REFERENCES

- Barron, O., O. Kim, S. Lim, and D. Stevens. 1998. Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review* 73 (October): 421–433.
- , C. Kile, and T. O'Keefe. 1999. MD&A quality as measured by the SEC and analysts' earnings forecasts. *Contemporary Accounting Research* 16: 75–109.
- Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics*. New York, NY: Wiley.
- Brown, L. 2001. How important is past analyst earnings forecast accuracy? *Financial Analysts Journal* 57 (6): 44–49.
- Bushee, B., D. Matsumoto, and G. Miller, 2001. Open versus closed conference calls: The determinants and effects of broadening access to disclosure. Working paper, University of Pennsylvania, University of Washington, and Harvard University.
- Francis, J., J. D. Hanna, and D. R. Philbrick. 1997. Management communications with security analysts. *Journal of Accounting and Economics* 24: 363–394.
- Frankel, R., M. Johnson, and D. Skinner. 1999. An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research* 37 (1): 133–150.
- Healy, P., A. Hutton, and K. Palepu. 1999. Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary Accounting Research* 16 (3): 485–520.
- Johnson, M., R. Kasznik, and K. Nelson. 2000. The impact of securities litigation reform on the disclosure of forward-looking information by high technology firms. *Journal of Accounting Research* 39 (2): 297–327.
- Kim, O., and R. Verrechia 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17: 41–67.
- Lang, M., and R. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 (October): 467–491.
- Lees, F. 1981. *Public Disclosure of Corporate Earnings Forecasts*. New York, NY: The Conference Board.
- Levitt, A. 1998. A question of integrity: Promoting investor confidence by fighting insider trading. Speech at SEC Speaks Conference, February 27. Available at: <http://www.sec.gov/news/speech/speecharchive/1998/spch202.txt>.
- McCafferty, J. 1997. Speaking of earnings. . . Why managing expectations often doesn't work. *CFO* 13 (10): 38–50.
- National Investor Relations Institute (NIRI). 2000. Updates on Issues Vital to the Practice of Investor Relations—Corporate Conference Calls. February 29. Vienna, VA: NIRI.
- . 1998a. Survey of corporate disclosure practices among U.S. companies. July. Vienna, VA: NIRI.
- . 1998b. Symposium on corporate disclosure: Impact of technology and role of media. April 8. Washington, D.C.: Government Printing Office.
- . 1998c. Utilizing technology in the practice of investor relations. April. Vienna, VA: NIRI.

- Opdyke, J., and M. Schroeder. 2001. Disclosure rule gets a bad rap. *Wall Street Journal* (June 5): C1.
- Securities and Exchange Commission (SEC). 2000. *Selective Disclosure and Insider Trading*. Proposed Rule S7-31-99. Available at: <http://www.sec.gov/rules/proposed/34-42259.htm>.
- Sinha, P., L. Brown, and S. Das. 1997. A re-examination of financial analysts' differential forecasting ability. *Contemporary Accounting Research* 14 (1): 1-42.
- Stickel, S. 1992. Reputation and performance among security analysts. *Journal of Finance* 47 (5): 1811-1836.
- Tasker, S. 1998. Bridging the information gap: Quarterly conference calls as a medium for voluntary disclosure. *Review of Accounting Studies* 3 (1-2): 137-167.

Materiality and Contingent Tax Liability Reporting

Cristi A. Gleason
Lillian F. Mills
The University of Arizona

ABSTRACT: We investigate factors that explain firms' decisions to disclose and record contingent tax liabilities. Our findings are based on confidential Internal Revenue Service audit data and financial statement footnotes for 100 large industrial firms from 1987 to 1995. Descriptive statistics indicate that these firms often fail to disclose IRS claims for tax deficiencies that exceed a 5-percent-of-income rule of thumb. We find that the probability of disclosure increases in the relative amount of the claim or the expected loss, although the largest claims drive this result. Our evidence is consistent with firms using a stable measure of size, such as assets or normal income, to gauge materiality, rather than relying only on current period reported income. We also find that the amount accrued for the contingent liability increases in the amount of the expected loss. However, our inferences may not generalize beyond a population of large, frequently audited firms.

Keywords: *materiality; discretionary disclosure; SFAS No. 5; tax cushion.*

Data Availability: *Firm financial statement data are available from public sources. Tax data are subject to disclosure protection under the Internal Revenue Code.*

I. INTRODUCTION

In his speech, "The Numbers Game," Securities and Exchange Commission (SEC) chairman Arthur Levitt (1998) decried the abuse of materiality as a "gimmick" firms use to manage earnings. Subsequently, the SEC (1999) issued Staff Accounting Bulletin

We appreciate helpful suggestions from Dan Dhaliwal, Leslie Eldenburg, Bill Felix, Sanjay Gupta, Don Hansen, Alister Hunt, Bill Kinney, Mary Margaret Myers, Mark Nelson, Kaye Newberry, John Robinson, Joel Slemrod, Kevin Smith, Steve Smith, Mark Trombley, and Hong Xie. This paper has benefited from comments by participants at the 2000 Arizona State University/University of Arizona Accounting Colloquium, the 2000 Mitsui Conference on Financial Economics and Accounting at the University of Michigan, and Oklahoma State University. Scott Odom and Ryan Weissmueller provided valuable research assistance. Professor Mills appreciates funding provided by the PricewaterhouseCoopers Tax Research Fellowship. Confidential tax return and audit data were obtained under a data non-disclosure agreement from the Internal Revenue Service, Large- and Mid-Size Business Division (LMSB), Strategy, Research & Program Planning Office, and are not available for use by others. All opinions are those of the authors solely and do not reflect views of the Internal Revenue Service.

Submitted February 2001
Accepted September 2001

(SAB) No. 99, *Materiality*. Lynn Turner (2000), SEC chief accountant, described SAB No. 99 as putting “people on notice that the use of simple quantitative cutoffs like 5 percent, or any other percent, as determining whether or not an item needed to be included or corrected are [sic] unacceptable. The real test is whether the information would make a difference when considered by a reasonable person.” Accounting for loss contingencies provides a setting where materiality judgments are complex but could make a difference to investors.

SFAS No. 5 (FASB 1975) requires that firms *disclose* an unaccrued loss if there is a reasonable possibility that the firm has incurred such a loss (§10); like all standards, it only applies to material items. Firms must *accrue* a loss contingency if it is probable and reasonably estimable (§ 8). Thus, the hurdle for disclosure (reasonably possible) is lower than the hurdle for recognition (probable and reasonably estimable). Assuming that the probability of a loss contingency is reasonably possible for our sample firms, we predict that the likelihood of disclosure increases in the amount, or materiality, of the Internal Revenue Service’s (IRS) claim that the firm has underpaid its federal income tax. We further predict that the amount the firm accrues for a probable loss contingency increases with the amount, or materiality, of the estimated expected loss.

SFAS No. 5 requires managers to exercise judgment about materiality and judgment about the probability of loss. Kieso and Weygandt (1998, 661) comment that applying SFAS No. 5 “involves judgment and subjectivity.... As a result, accruals and disclosures of contingencies vary considerably in practice.” Thus, financial statement users are uncertain whether and how firms will disclose loss contingencies. Evidence on factors explaining firms’ decisions to disclose and accrue contingent tax liabilities can help financial statement users interpret the presence or absence of disclosure to make inferences about possible or probable losses.

We examine contingent tax liabilities as a specific case of contingent losses. We use confidential Internal Revenue Service (IRS) audit data (subject to disclosure protection under the Internal Revenue Code) on the potential liability a firm faces for underpaying its federal income taxes. Several features of our sample and data make tax contingencies a powerful context in which to study disclosure and recording of contingent liabilities. The firms in our sample are part of the Coordinated Industry Cases (formerly known as the Coordinated Examination Program, also called the large-case audit program) of the IRS.¹ Because the IRS continually audits these firms, a loss contingency exists for each year. Although sophisticated investors may know that the IRS is likely auditing the company, investors are not privy to the firm’s private information about the likelihood that the IRS will prevail in a claim for additional tax revenue. We provide evidence on the extent to which firms follow SFAS No. 5 requirements in a setting in which investors lack value-relevant information about the contingent tax liability, unlike the environmental liability setting in which investors can obtain Superfund data.

We read financial statement footnotes for a sample of 100 large manufacturing firms over a nine-year period to obtain and classify contingent loss disclosures. The firms in our sample disclose very little about tax audits. Only 27 percent of firms make any disclosure

¹ Although the IRS seldom audits small corporations, it audits the largest firms nearly every year. More than 1,000 of the largest U.S. corporations are part of the IRS’ Coordinated Industry Cases. Firms are included in the Coordinated Industry Cases according to a point system that measures the size and complexity of the corporate tax return, as described in the *Internal Revenue Manual*. The IRS coordinates the examination among team members with specialized knowledge in the industry, international taxation matters, engineering, economics, and employment and excise taxes. These firms account for more than 50 percent of the corporate tax revenues in the United States (IRS 1995).

of contingent tax liabilities and only 30 percent of firms that do disclose a tax contingency provide the detailed information required by SFAS No. 5.

Using footnote data on loss disclosure and IRS audit data on tax deficiencies, we estimate a logit regression to explain management's disclosure decisions. We find that the likelihood that the firm will disclose the contingent tax liability increases in the amount of either the claim or the expected loss, measured as the claim multiplied by the industry-level settlement percentage (collections/claims for settled returns). This result is primarily attributable to the largest claims, consistent with our anecdotal understanding of how firms and auditors apply SFAS No. 5 in practice. Audit and tax partners that we interviewed indicate that the expected loss is the primary consideration in disclosure decisions, but extremely large claims and public awareness of the claim via press reports both trigger disclosure.

Our results suggest that firms gauge materiality relative to a stable measure of size, such as "normal income" (the greater of income or 5 percent of assets) or assets, rather than by current period reported income alone. We control for temporary differences between book and tax income and other factors associated with discretionary disclosure, such as firm size, new capital issuance, susceptibility to litigation, and disclosure quality.² The results show that firms operating in litigious industries are more likely to disclose contingent tax liabilities, consistent with Skinner's (1994) conclusion that disclosure reduces the cost of potential litigation. Because our sample comprises large, frequently audited firms, our results may not generalize to all corporations.

We also estimate an ordinary least squares (OLS) regression to test whether the expected loss explains the amount of the accrual for the loss contingency. We estimate the accrual as the accumulated difference between U.S. current tax expense and the tax on the return for all open tax years. We estimate the expected loss as the claim multiplied by the industry-level settlement percentage (collections/claims for settled returns). The accrual for the contingent tax liability increases with the expected loss, controlling for temporary differences between book and tax accounting and other variables associated with discretionary disclosure.

Our finding that although firms disclose the largest claims, they do not disclose many so-called "material" claims, suggests that many firms do not use the oft-cited 5-percent-of-income materiality benchmark as academics and regulators may expect. Furthermore, prior research does not show how a firm gauges materiality when it has a loss or a low-income year. Our evidence suggests that materiality is based on "normal income" when the firm reports low income or a loss. If firms only disclose contingent losses when they are material compared to "normal income," then shareholders may misinterpret the lack of disclosure as an absence of material contingencies, especially in the first year of low income.

II. PRIOR LITERATURE AND HYPOTHESES DEVELOPMENT

Materiality

When the SEC identified materiality as an earnings management tool, researchers focused increasing attention on materiality (i.e., Libby and Kinney 2000; Nelson et al. 2001). Prior research consistently finds that the effect of an adjustment on income is the primary factor influencing an auditor's decision to qualify the audit opinion (Moriarty and Barron 1976; Bates et al. 1982; Chewning et al. 1989). For example, Chewning et al. (1989) find

² See, for example, Lang and Lundholm (1993, 1996), Skinner (1994, 1997), Frankel et al. (1995), Botosan (1997), Francis et al. (1998), Ali and Kallapur (2001), Miller and Piotroski (2000), Botosan and Plumlee (2001).

that few auditors issue unqualified audit reports for firms making accounting changes that exceed 4 percent of income. In contrast to prior research that focused on the audit opinion, we provide empirical evidence on the materiality of disclosed contingent liabilities.

Prior research does not consider how a firm evaluates income-based materiality when it has a loss or a low-income year. We consider the following alternative benchmarks to assess materiality: (1) total assets, (2) income for profit firms (including low-income firms) and 5 percent of assets for loss firms, and (3) the greater of income or 5 percent of assets, which we interpret as a measure of "normal income." We provide evidence on which of these three measures best reflects firms' materiality decisions as applied to contingent liability reporting.

Contingent Liabilities

Our paper also extends research on contingent liability accounting. Empirical research on contingent liabilities has focused primarily on the relation between security prices and contingent liability disclosure.³ Banks and Kinney (1982) and Frost (1991) find a significantly negative price reaction when a firm reports new contingent liabilities, suggesting that contingency footnotes provide value-relevant information. Using Superfund site data, Barth and McNichols (1994) find that their estimate of a firm's environmental liability is associated with share prices. More directly related to our study, Barth et al. (1997) examine the factors affecting whether and in what detail a firm discloses contingent environmental liabilities (see also Li et al. 1997). Because Superfund site-related environmental costs are large and well publicized, and additional standards provide more guidance on when firms must disclose environmental liabilities, firms frequently disclose these costs. Thus, Barth et al. (1997) perform a rich study of the nature of that disclosure. Conversely, our setting informs researchers and financial statement users about firms' disclosure of contingent liabilities where shareholders lack external data about the existence, amount, or nature of the loss contingency.

Discretionary Disclosure

We expect that the primary explanation for disclosure will be the SFAS No. 5 mandatory requirement to disclose material loss contingencies. However, the discretion that management exercises in applying the disclosure rules means that the propensity to issue contingent tax disclosures is likely related to other factors associated with discretionary disclosure, including size, capital issuance, likelihood of litigation, and overall disclosure quality. For example, Lang and Lundholm (1993, 1996), Frankel et al. (1995), Botosan (1997), and Botosan and Plumlee (2001) conclude that large firms make more frequent voluntary disclosures, and that firms with better overall disclosure enjoy a lower cost of capital. Because firms seeking new capital are especially sensitive to the cost of capital, we control for new capital issuance as well as for firm size and overall disclosure quality.

Skinner (1994, 1997) finds that firms preemptively disclose bad news more often than good news. Disclosing bad news does not reduce the likelihood of litigation, but it may reduce the cost of resolving any resulting litigation. We consider whether the cost of potential litigation motivates firms to disclose tax contingencies by comparing firms in litigious industries (Francis et al. 1998; Ali and Kallapur 2001; Miller and Piotroski 2000) with firms not in these industries.

³ A substantial literature exists on loan loss provisions, a material contingency for banks and insurance firms (see for example, Petroni 1992; Wahlen 1994; Liu et al. 1997). Banks and insurance firms disclose both the existence of the contingency and the accrued liability. In contrast, our research setting involves external tax liability claims that shareholders may not know about, where disclosure is infrequent.

Disclosure Hypothesis

We first examine the nature of tax contingency disclosures in the financial statement footnotes. A firm must disclose a material contingent liability in a footnote to the financial statements if a loss is probable or reasonably possible (SFAS No. 5). We assume that a loss is probable or reasonably possible when the IRS proposes a deficiency, because our sample firms settle 70 percent of proposed tax deficiencies for a positive amount.⁴

SFAS No. 5 does not clearly specify whether firms should apply the materiality threshold to the original claim or to the amount the firm expects to lose. SAB No. 92 (SEC 1993), which applies to environmental liabilities, cautions “that a statement that the contingency is not expected to be material does not satisfy the requirements of SFAS No. 5 if there is a reasonable possibility that a loss exceeding amounts already recognized may have been incurred and the amount of that additional loss would be material.” The spirit of these rules could indicate that standard setters believe firms should evaluate the materiality of the claim in addition to the expected loss. However, interviews with audit and tax partners of Big 5 CPA firms indicated that they do not apply environmental contingency standards such as SAB No. 92 and Statement of Position No. 96-1, *Environmental Remediation Liabilities* (AICPA 1996), to other types of contingencies. Practitioners also revealed they consider the materiality of the expected loss rather than the materiality of the claim, unless the claim is especially large. We predict that the likelihood that the firm discloses contingent tax liabilities increases with the amount of the *claim* (consistent with a conservative reading of SFAS No. 5) or the amount of the *expected loss* (consistent with practitioners’ views), in the following alternative hypothesis:

H1: The more material the claim (or expected loss), the more likely the firm will disclose a contingent tax liability in the footnotes.

Recording Hypothesis

Following SFAS No. 5, firms should record contingent liabilities that are both probable and estimable. Tax contingency claims rarely relate to a single item that the firm will either win or lose in total, but consist of numerous items of disagreement. We assume that the requirements of “probable and estimable” are intertwined such that the manager uses his beliefs about the probability of loss of individual items to estimate the total contingent loss, consistent with our interviews with tax partners.

Prior evidence about accruals for tax contingencies, commonly called the “tax cushion,” is largely limited to a brief reference that the tax cushion may explain the discrepancy between current tax expense and the tax liability (see Dworin 1985, note 9). Smith (2001) conducts an experimental survey to determine when tax managers record tax cushions and finds that they are more likely to record tax cushions for deductions that generate permanent book-tax differences. We extend this experimental research by providing empirical, archival evidence about whether firms record larger tax cushions when they expect greater losses:

H2: The amount the firm accrues as a contingent liability (tax cushion) increases with the amount of the expected loss.

⁴ In the Coordinated Industry Cases population as a whole during the period 1981 to 1996, out of 25,820 firm-year observations, 91 percent were subject to a full audit. Of 16,690 firm-year observations with audits completed, 55 percent had a positive audit adjustment. Of 12,395 firm-year observations with closed years (the audit is no longer under dispute), 59 percent settled for a positive amount. In our sample, 94 percent of firm-years are audited, 71 percent of audited firms have a positive audit adjustment, and 70 percent of closed years settle for a positive amount.

Our two hypotheses imply that managers make disclosure and accrual decisions independently of one another. Sensitivity tests (reported later) reveal no evidence that firms make these decisions simultaneously. Moreover, audit and tax partners that we interviewed stated that the disclosure and the accrual decision take place at different times and are made by somewhat different personnel. Specifically, firms record some tax cushion in the year they file the tax return and they update this accrual as the audit, appeals, and litigation processes reveal more information. Audit partners state that firms do not consider disclosing tax contingencies until the IRS assesses the claim, and that firms that have already recorded an adequate provision for the expected loss are less likely to disclose. We therefore test the two hypotheses independently, and our disclosure test controls for the adequacy of the accrued tax cushion.

III. VARIABLE DEFINITIONS AND ESTIMATION EQUATIONS

Measuring the Contingent Liability

Our sample firms typically have contingencies relating to several tax-return-years because of the length of time required for the IRS to complete the audit and for the two parties to settle any resulting disputes. We measure the claim (\$CLAIM) as the cumulative amount of unresolved IRS tax deficiencies from 1981 (the first year for which IRS audit data are available) to three years prior to the disclosure year.⁵ For example, for 1990, we include all unresolved claims from 1981 to 1987. We end our accumulation three years before the disclosure year because three years is the length of the statute of limitations (under Internal Revenue Code Section 6501). Furthermore, discussions with the IRS indicate that three years generally represents a minimum length of time to complete an audit and propose any deficiencies for this sample. We believe our cumulative measure is conservative. If the IRS first notifies a firm about a claim more than three years after the firm files the return, then our measure biases against finding a significant correlation between claims and disclosure.⁶

In determining the expected loss, firms must estimate the probability associated with the possible outcomes of the items comprising the claim. Firms have historical information about their success in previous disputes. They may also have information about other firms in their industries or about firms with similar disputes, either through professional associations such as the Tax Executives Institute or through their certified public accounting or law firms. We use the average settlement rate for all settled claims in the industry (using IRS industry codes) to proxy for the firm's estimate of probable loss because we do not have enough closed return observations per firm to estimate firm-specific settlement rates.

⁵ Ideally, we would like to measure the information regarding the claim that the IRS reveals to the firm during the disclosure year. However, the IRS data do not identify when an audit began or when the IRS proposed a deficiency. We know the year of the tax return and three other dates related to the settlement of the audit: the latest date upon which an Exam issue was agreed, or the dates upon which Appeals or Counsel made an assessment. We use the latest of these three dates to identify the year the parties resolve the claim. We ignore all unaudited returns or incomplete audits, because data are not available to estimate the loss contingency, and because firms may consider claims on unaudited returns to be too remote to require disclosure.

⁶ We can view the claim as a noisy measure of tax aggressiveness. If tax-aggressive firms fight the claims more vigorously, and more successfully, then such firms may assign a lower probability of IRS success to a given amount of claim. The lower probability will result in a smaller likelihood of disclosure, and a smaller amount of cushion accrued. Thus, tax aggressiveness biases against detecting the predicted relation in H1 and H2. We compare the firm-specific settlement percentage (collection/claim) for all settled returns across the deciles of the claim (scaled by normal income). We find neither that the settlement percentage declines monotonically as claims grow nor that the settlement percentage is lower for the top claim deciles. Thus, we do not have evidence that tax aggressive firms are more successful at settling for a lower percentage of the claim than are firms with smaller claims.

FASB Interpretation (FIN) No. 14 states that when the firm estimates that a loss will fall within a range, but no amount within the range is a better estimate of the loss than another amount, the firm should accrue the minimum amount (FASB 1976). If a firm bases its accrual on a lower bound of a range, then cushion will be less than 100 percent of our measure of expected loss.

Tax Cushion

We measure the annual amount of the tax cushion as the difference between U.S. current income tax expense reported in the financial statements and total tax on the income tax return.⁷ Consistent with our cumulative measure of the claim, CUSHION is the cumulative amount accrued in years with unresolved tax claims.⁸ Although Dworin (1985) indicates that the cushion is recorded in the current tax expense, SFAS No. 109 (FASB 1992) provides little guidance about how the tax cushion relates to the current vs. deferred components of income tax expense. However, because firms must eventually record obligations for payment to the IRS as current liabilities, our cumulative measure of current tax expense less tax on the return should capture the total cushion for open tax years.⁹

Measuring Materiality

We use a proxy for “normal income” as a reference point to evaluate materiality. Normal income provides a more stable basis for loss firms and for firms with abnormally low income than does reported net income. Leslie (1985, 30) analyzes cases reported by Pattillo (1976) and concludes that “materiality decisions for loss and break-even firms were within 5% to 10% of what might be considered normal net income.” The median pretax return on assets for our sample is 9 percent; thus 5 percent of assets is a conservative estimate of normal income. We define %CLAIM as the total claim (\$CLAIM) divided by the maximum of pretax income or 5 percent of total assets. In supplemental tests, we find that inferences are identical if we scale \$CLAIM by (1) the greater of pretax income or 10 percent of assets; or (2) total assets. However, results are sensitive to using 5 percent of income as a scalar only for loss firms.

Controlling for Balance Sheet Effects

Finally, we consider how the unique nature of a tax contingency can affect the income statement and the balance sheet differently. In the typical accrual for a contingent loss, the debit to income statement expense equals the credit to the balance sheet liability. However, tax deficiencies affect earnings only when they relate to permanent items such as permanent income differences, credits, interest, and penalties. Deficiencies related to temporary differences create balance sheet reclassifications (although the interest due on deficiencies related to temporary differences is an expense). For example, if the IRS requires a firm to capitalize and depreciate a previously expensed purchase, then earnings is not affected but

⁷ Where U.S. current tax expense is missing (70 observations), we use total tax expense (Compustat item no. 16) less total deferred tax expense (item no. 50) if available (for 41 observations), or total tax expense (for 29 observations). Our tests are robust to estimating U.S. current tax expense as total current tax expense multiplied by the ratio of U.S. pretax income (item no. 272) if available (for 61 observations), divided by consolidated pretax income (item no. 170).

⁸ Where the accumulated cushion (CUSHION) is less than zero, we set it to zero, because we want to proxy for the amount of aggregate accrual for current tax in excess of tax recorded on the tax return. This censoring affects 195 firm-year observations. The results are unchanged if we (1) restrict the sample to the 705 observations with positive cushion, or (2) permit the cushion to be negative.

⁹ See Hanlon and Shevlin (2001) regarding how stock option deductions reduce tax on the return but not book tax expense, potentially adding noise to our measure.

the firm must pay taxes sooner. Assuming a fixed standard (such as a 5 percent rule-of-thumb), the absolute materiality threshold is lower for earnings adjustments (that are assessed relative to normal income) than for balance sheet reclassifications that are assessed relative to total assets (Leslie et al. 1980; Leslie 1985). Thus, we expect that firms are less likely to disclose deficiencies due to temporary differences reflected in balance sheet reclassifications than to disclose deficiencies due to permanent differences that affect earnings.

We control for the portion of the contingent tax liability that affects only the balance sheet using a proxy for the amount of the audit adjustment that is related to temporary differences. Prior research argues that book-tax differences represent aggressive tax positions (Cloyd 1995; Cloyd et al. 1996; Mills 1998). Consistent with this argument, Mills (1998) finds that proposed tax deficiencies are positively related to both (1) total differences between book income and taxable income and (2) U.S. deferred tax expense (a proxy for temporary differences between book and tax income). Thus, we use the U.S. deferred tax expense (USDEF_TAX), accumulated for all open tax return years, to control for the deficiencies related to temporary differences. In combination with %CLAIM, this provides a more complete picture of the materiality of the tax contingency, because claims that affect only the balance sheet are less likely to be material.

Test of H1

To investigate the relation between the firm's decision to disclose a contingent tax liability and the amount of the claim (or expected loss), we estimate a logistic regression of a binary variable (DISCLOSE = 1 or 0) on the amount of the claim (or expected loss) and other control variables that predict disclosure. Unlike Barth et al. (1997), who use the wide range of disclosure content to create a pseudo-continuous dependent variable for OLS regression, we use a binary disclose/nondisclose variable because the tax disclosures are infrequent and often thin. Our estimation equation is as follows:

$$\ln \frac{P_{disc}}{1 - P_{disc}} = \alpha + \beta X + \varepsilon$$

where $P_{disc} = \frac{1}{1 + e^{-(\alpha + \beta X + \varepsilon)}} = \text{probability the firm will disclose a contingent tax liability.}$

$$\begin{aligned} \beta x = & \beta_1 \%CLAIM_{i,t} \text{ (or EXP_LOSS}_{i,t}) + \beta_2 ADEQUATE_{i,t} + \beta_3 USDEF\_TAX_{i,t} \\ & + \beta_4 LOGASSETS_{i,t} + \beta_5 NEWCAP_{i,t} + \beta_6 LITIGIOUS_{i,t} \\ & + (\beta_7 DISCLOS\_QUAL_{i,t}) + \varepsilon_{i,t} \end{aligned} \quad (1)$$

where:

DISCLOSE = 1 if the firm discloses any information with respect to an IRS audit, and 0 otherwise;

%CLAIM = cumulative proposed tax deficiencies for unresolved tax claims since 1981 through three years prior to the disclosure year (\$CLAIM), scaled by the maximum of total pretax income (Compustat item no. 170) or 5 percent of total assets (Compustat item no. 6);

- EXP_LOSS = %CLAIM times the average settlement rate for the industry (using IRS industry codes);
- ADEQUATE = 1 if CUSHION > EXP_LOSS, 0 otherwise. CUSHION equals the cumulative difference between U.S. current income tax expense reported in the financial statements and total tax on the income tax return in unresolved years, scaled by the maximum of total pretax income or 5 percent of total assets;
- USDEF_TAX = cumulative amount of U.S. deferred income tax expense (Compustat item no. 269) for all unresolved tax years since 1983 (the first year available on Compustat) through the disclosure year, scaled by the maximum of total pretax income or 5 percent of total assets;
- LOGASSETS = the natural log of total assets (Compustat item no. 6);
- NEWCAP = 1 if the firm issued new debt or equity during the year according to the Securities Data Corporation, 0 otherwise;
- LITIGIOUS = 1 if the firm is in a litigious industry (Francis et al. 1998; Ali and Kallapur 2001; Miller and Piotroski 2000), 0 otherwise (where SIC 2834, 3570, 3571, 3577, 3663, and 3674 are litigious industries in our sample); and
- DISCLOS_QUAL = normalized total score of disclosure quality by the Financial Analysts' Federation (FAF), used in a separate model for the subsample of firms for which DISCLOS_QUAL is available.

Hypothesis 1 posits that the probability that the firm makes a disclosure about its contingent tax liability is positively related to %CLAIM, even after controlling for other factors that are likely to affect disclosure. Based on discussions with practitioners, we also expect the probability of disclosure to be positively related to EXP_LOSS.

In addition to the magnitude of the claim or expected loss, we expect that the probability that the firm discloses a contingent tax liability depends on other factors associated with the contingent tax liability specifically and with voluntary disclosure in general. We expect that firms with adequate cushion accrued are less likely to disclose a loss contingency, so we expect the coefficient on ADEQUATE to be negative. As the amount of the claim relating to balance sheet reclassifications rather than net income increases, the likelihood that the claim is material to the income statement decreases. Therefore, we expect the coefficient on USDEF_TAX to be negative. LOGASSETS controls for the effects of size in our scaled explanatory variables. Although Botosan (1997) finds that larger firms make more complete financial disclosures, all of our sample firms are extremely large. Thus, firm size may not explain much variation in our sample firms' disclosure.

We also include three additional controls for discretionary disclosure. We include NEWCAP, a dummy variable that equals 1 if the firm issued new debt or equity during the year, and 0 otherwise. If firms are reluctant to disclose contingent liabilities during a year in which they obtain new capital, then the coefficient on NEWCAP will be negative. We also include LITIGIOUS, which equals 1 if the firm is in a litigious industry (Francis et al. 1998; Ali and Kallapur 2001; Miller and Piotroski, 2000), 0 otherwise. We expect that firms in litigious industries are more likely to disclose contingencies.

Finally, we include DISCLOS_QUAL, which indicates how the Financial Analysts Federation (FAF) Corporate Information Committee ranks the firm's overall quality of disclosure (within industry), using the normalized total score which is a weighted combination

of scores for annual published information, quarterly and other published information, and investor relations (see Lang and Lundholm 1993, 1996). We expect that firms that on average provide high-quality disclosures are more likely to disclose contingent tax liabilities.

Test of H2

We use a similar equation to investigate the relation between the magnitude of the tax cushion accrued and the magnitude of the expected loss as H2 predicts:

$$\text{CUSHION}_{i,t} = \beta_0 + \beta_1 \text{EXP\_LOSS}_{i,t} + \beta_2 \text{USDEF\_TAX}_{i,t} + \beta_3 \text{LOGASSETS}_{i,t} + \beta_4 \text{NEWCAP}_{i,t} + \beta_5 \text{LITIGIOUS}_{i,t} + (\beta_6 \text{DISCLOS\_QUAL}_{i,t}) + \varepsilon_{i,t}, \quad (2)$$

where variables are defined as for Equation (1).

Hypothesis 2 posits that the coefficient for EXP_LOSS is positive. We expect USDEF_TAX to be negative because expected tax payments due to temporary differences do not require earnings accruals. NEWCAP will have a negative coefficient if firms that issue new debt or equity are reluctant to accrue losses. LITIGIOUS will have a positive coefficient if firms that are at higher risk of lawsuits are more conservative with respect to accruing losses. DISCLOS_QUAL will also have a positive coefficient if firms with high-quality disclosure are more conservative. Supplemental tests (described later) reveal no endogeneity between the decision to disclose and the amount of the tax cushion accrued, so we estimate the two equations separately.

IV. DATA AND EMPIRICAL RESULTS

Data and Descriptive Statistics

Our sample comprises the largest industrial firms in the intersection of the Compustat PST file and the database of IRS audit and tax return data for the Coordinated Industry Cases. The largest firms comprise a sample for which the high frequency of audits and proposed tax deficiencies suggest that a loss contingency is more than remote. If large firms are more conservative in tax reporting, then our sample would conservatively measure potential underreporting in the general population. Because large corporations can use more sophisticated tax planning to develop creative tax positions, we do not generalize our results to the population of smaller corporations, which are infrequently audited.

We began with an intersection of the 200 largest manufacturing firms (by 1995 assets, data item no. 6) on the Compustat PST tape and in the IRS Coordinated Industry Cases data. This resulted in a sample of 117 unique firms. The final sample is 100 firms with tax return and audit data for all 15 years between 1981 and 1995 and nine years of tax and contingency footnote disclosures from 1987 through 1995. The relevant IRS audit data, obtained under a data disclosure agreement, include the tax paid on the return, the claim, and the collection on the claim, if any.

Table 1, Panel A presents descriptive statistics. Only 27 percent of the firm-years include a disclosure about tax audits. CUSHION is approximately 39 percent of normal income on average, although the median is only 23 percent. The claim (\$CLAIM) is \$302 million, on average. We omit the maximum value to preserve confidentiality, but the maximum claim exceeds \$1 billion. Claims average nearly 20 percent of normal income (the larger of pretax income or 5 percent of assets). However, the median claim is only 6 percent of income. The expected loss is far smaller with a mean (median) of 5 percent (2 percent) of normal income. U.S. deferred tax expense for open tax years accumulates to 4

TABLE 1
Descriptive Statistics for Dependent Variables (Disclosure and Accrued Cushion for Tax Contingencies), Explanatory Variables (Various Measures of the IRS Claim), Control Variables, and the Financial Statement vs. Tax Return Comparison Variables

Panel A: Disclosure, Audit, Tax, and Financial Variables for Pooled Sample						
Variable ^a	<i>n</i> ^b	Mean	Std. Dev.	Minimum	Median	Q3
DISCLOSE	900	0.27	0.45	0	0	1
CUSHION	900	0.39	0.48	0	0.23	0.46
\$CLAIM	900	301.54	965.11	0	52.21	186.77
%CLAIM	900	0.19	0.37	0	0.06	0.20
CLAIMASST	900	0.02	0.03	0	0.01	0.02
EXP_LOSS	900	0.05	0.11	0	0.02	0.06
ADEQUATE	900	0.74	0.44	0	1	1
USDEF_TAX	900	0.04	0.43	-3.36	0.06	0.23
LOGASSETS	900	9.04	0.99	6.80	8.81	9.46
NEWCAP	900	0.37	0.48	0	0	1
LITIGIOUS	900	0.12	0.32	0	0	0
DISCLOS_QUAL	477	0.59	0.30	0.06	0.63	0.86

(Continued on next page)

TABLE 1 (Continued)

Panel B: Comparison of Financial Statement vs. Tax Return Variables for Purposes of Understanding CUSHION Estimation

	Financial Statements		Tax Return		<i>t</i> -test of Differences
	Definition	Mean (n)	Definition	Mean (n)	
Income	U.S. pretax income (item no. 272)	598.49 (n = 767)	Taxable income (return line 28)	869.59 (n = 888)	-4.13***
	Worldwide pretax income (item no. 170)	1,134.62 (n = 900)	Taxable income (return line 28)	869.59 (n = 888)	
	U.S. current tax expense (item no. 63)	182.05 (n = 830)	Total tax (return line 31)	146.80 (n = 900)	
Assets	Total assets (item no. 6)	16,562.40 (n = 900)	Total assets (return Schedule L)	18,372.26 (n = 888)	1.20
Tax					3.05***

^a Variable definitions:

DISCLOSE = 1 when the firm discloses a contingent tax liability. Specifically, we code Disclosure as 1 if it is a DISPUTE or OPENYRS or MENTION (see Appendix), 0 otherwise. CUSHION is the accrued tax contingency, estimated as the accumulated difference between U.S. current tax expense from Compustat (item no. 63) and total tax on the U.S. corporation income tax return (Form 1120, line 31) for all open tax years (still under Audit, Appeal, or Counsel), scaled by normal income. \$CLAIM is the cumulative proposed tax deficiency for open tax years (still under Audit, Appeal, or Counsel). %CLAIM is \$CLAIM scaled by normal income (the greater of total pretax income [Compustat item no. 170] or 5 percent of assets [Compustat item no. 6]). CLAIMASST is \$CLAIM scaled by assets (Compustat item no. 6). EXP_LOSS is the cumulative expected loss, estimated as the cumulative proposed tax deficiency for open tax years multiplied by the average industry settlement ratio (payments to the IRS/proposed tax deficiencies), scaled by normal income. ADEQUATE = 1 when CUSHION exceeds EXP_LOSS, 0 otherwise. USDEF_TAX is the cumulative U.S. deferred income tax expense (Compustat item no. 63) for open tax years, scaled by normal income. LOGASSETS is the natural log of assets on Compustat (item no. 6). NEWCAP = 1 when the firm issued debt or equity in the current year according to the Securities Data Corporation, 0 otherwise. LITIGIOUS = 1 if the firm is in a litigious industry (Francis et al. 1998; Ali and Kallapur 2001; Miller and Piotroski 2000), 0 otherwise. DISCLOS_QUAL is the firm's industry-adjusted rank of disclosure quality according to the FAF.

^b The sample consists of nine years of data from 1987 to 1995 for the 100 largest industrial firms in the intersection of the Compustat PST file and the database of IRS audit and tax return data for the Coordinated Industry Cases (large-case audit firms). The relevant tax audit data, obtained under a data disclosure agreement subject to disclosure protection under the Internal Revenue Code, show the tax paid on the return, the amount of the tax deficiency, and the amount collected on the claim, if any.

percent of income on average, although the sample includes both positive and negative balances of accumulated deferred tax expense. Sample firms are quite large, as expected, with a mean log of assets of \$9 million, which is equivalent to \$8 billion of total assets. Sample firms issued public debt or equity (NEWCAP) in approximately 37 percent of firm-years. About 12 percent of our firms are in LITIGIOUS industries. For those 477 firm-years for which the FAF ranks disclosure quality, the average normalized rank of 59 percent indicates that sample firms have higher disclosure quality than other firms in their industry.

Panel B of Table 1 compares information from the financial statements to the tax return. Worldwide pretax income exceeds taxable income, and U.S. pretax income is less than taxable income. Mills and Newberry (2001) explain that U.S. taxable income includes repatriations of foreign-jurisdiction income (such as dividends) that firms would not report as U.S. pretax income on Compustat. However, because total tax on the return is net of foreign tax credits, U.S. current tax expense and total taxes are approximately jurisdictionally equivalent. U.S. current tax expense exceeds total tax on the return on average, and this difference is our estimate of CUSHION. Although total assets for book and tax purposes can have some differences in consolidation, the means are not statistically different.

We collect footnote disclosures for fiscal years from 1987 to 1995 to determine when and what firms disclose about tax contingencies. Forty-one percent of firms in our sample make some disclosure about IRS tax audits at least once during the period. We classify the types of disclosure using four categories (see the Appendix for examples of typical disclosures):

- (1) No disclosure (NONE).
- (2) Firms that mention a tax contingency only as part of a group of possible contingencies (MENTION). This disclosure typically appears in the Commitments and Contingencies footnote.
- (3) Firms that disclose the status of audits for prior tax returns (OPENYRS). Some firms report the years open and under audit. Others report the most recent year, typically several years earlier, for which the audit is complete and all disputes settled.
- (4) Firms that disclose a notice of deficiency or adjustment (DISPUTE). This final category represents the most disclosure and is the only one in which firms provide the information specified in SFAS No. 5: the nature of the dispute, the issue under appeal, the amount of the claim, and whether the firm has booked an accrual for possible losses.

Table 2, Panel A reports the summary statistics by disclosure category and provides univariate evidence that disclosure increases with the materiality of the contingent tax liability. Most of the disclosures about contingent tax liabilities provide no information about the nature of the dispute or the amounts involved. About 20 percent of firms that disclose a contingent tax liability MENTION the tax contingency among other contingencies. Approximately 50 percent of firms disclosing a contingent tax liability disclose OPENYRS. Only 30 percent of tax contingency disclosures are categorized as DISPUTE. We compare the means of the unscaled and scaled claim and expected loss (\$CLAIM, %CLAIM, CLAIMASST, EXP_LOSS) across the four categories of disclosure. Firms that disclose a DISPUTE with the IRS have significantly higher claims and EXP_LOSS than any other disclosure category. Furthermore, firms that MENTION a tax contingency have higher claims than do firms that make no disclosure. There is no significant difference between firms that disclose nothing and firms that disclose OPENYRS.

Panel B of Table 2 shows that the frequency of each type of disclosure appears relatively stable over time, although firms make more open year disclosures over the sample period.

TABLE 2
Descriptive Statistics for Disclosure of Tax Contingencies and IRS Claims over the 1987 to 1995 Sample Period

Panel A: t-tests of How Mean Claim Differs by Disclosure Class

<i>Variables^a</i>	<i>Type of Disclosure</i>			<i>t-tests Indicate Means Are Not Significantly Different at the 5% Level</i>
	<i>NONE</i>	<i>MENTION</i>	<i>OPENYRS</i>	
# observations	655	49	123	73
\$CLAIM	208.33	424.78	132.07	1,353.66
%CLAIM	0.13	0.25	0.16	0.74
CLAIMASST	0.01	0.02	0.01	0.06
EXP_LOSS	0.05	0.08	0.06	0.16
				NONE = OPENYRS
				MENTION = OPENYRS, NONE = OPENYRS
				MENTION = OPENYRS, NONE = OPENYRS
				MENTION = OPENYRS = NONE

Panel B: Annual Disclosure and Mean Claim

<i>Year</i>	<i>n</i>	<i>Type of Disclosure</i>			<i>Mean \$CLAIM</i>	<i>Mean %CLAIM</i>
		<i>NONE</i>	<i>MENTION</i>	<i>OPENYRS</i>		
1987	100	77	4	11	194.12	0.125
1988	100	76	4	12	225.98	0.130
1989	100	77	3	12	254.44	0.146
1990	100	76	4	12	301.93	0.162
1991	100	74	6	12	331.42	0.227
1992	100	74	6	13	340.92	0.236
1993	100	69	8	15	357.15	0.256
1994	100	67	7	17	356.86	0.225
1995	100	65	7	19	360.53	0.195

Panel C: Frequency of Total Number of Years of Disclosure by Each Sample Firm

Number of years the firm discloses during the nine sample years	0	1	2	3	4	5	6	7	8	9
Number of firms disclosing this many times:	59	4	3	4	3	5	2	2	2	16

^a See Table 1 for variable definitions and sample construction

In addition, we report the annual mean claim. Because we aggregate proposed tax deficiencies from 1981 until three years prior to the disclosure year, the later years in the sample show that the claim grows over the sample period. However, the scaled claim levels out in the middle of the period.¹⁰ Panel C reports that 16 companies disclose tax contingencies in all nine years of the 1987 to 1995 sample period. Eleven other firms make disclosures in more than half of the years.

Figure 1 displays the association between firm disclosure and information about the contingent liability. In Panel A, we form deciles by %CLAIM. We plot the percentage of firms in the decile that disclose a tax contingency and the median value per decile for %CLAIM and USDEF_TAX. We observe that disclosure substantially increases only in the 10th decile, in which the median claim exceeds 80 percent of normal income. In the first nine deciles, only 20 to 30 percent of firms disclose a tax contingency. In Panel B we form deciles by EXP_LOSS, and plot the percentage disclosing and the median EXP_LOSS and USDEF_TAX. Disclosure does not increase smoothly across the deciles of EXP_LOSS, although the 10th decile shows the highest (40 percent) disclosure.

If all of the claim affected earnings, then we would conclude that disclosure was inadequate, because only about 20 percent of firms disclose a claim, even when the claim exceeds 5 percent of normal income in the fifth decile. However, tax contingencies can affect both earnings and balance sheet items. The relation between disclosure and temporary differences (USDEF_TAX) is consistent with our expectations in both panels. Disclosure generally moves opposite deferred tax expense, with the exception of the 10th decile of Panel A, where the largest claims are disclosed even if they are related to temporary differences.

Table 3 shows the means of our explanatory and control variables by DISCLOSE and the correlations between cushion (CUSHION) and our explanatory and control variables. All of our measures of claim are significantly larger for firms that disclose a tax contingency (DISCLOSE = 1), consistent with H1. CUSHION is significantly positively correlated with the expected loss (EXP_LOSS), consistent with H2.

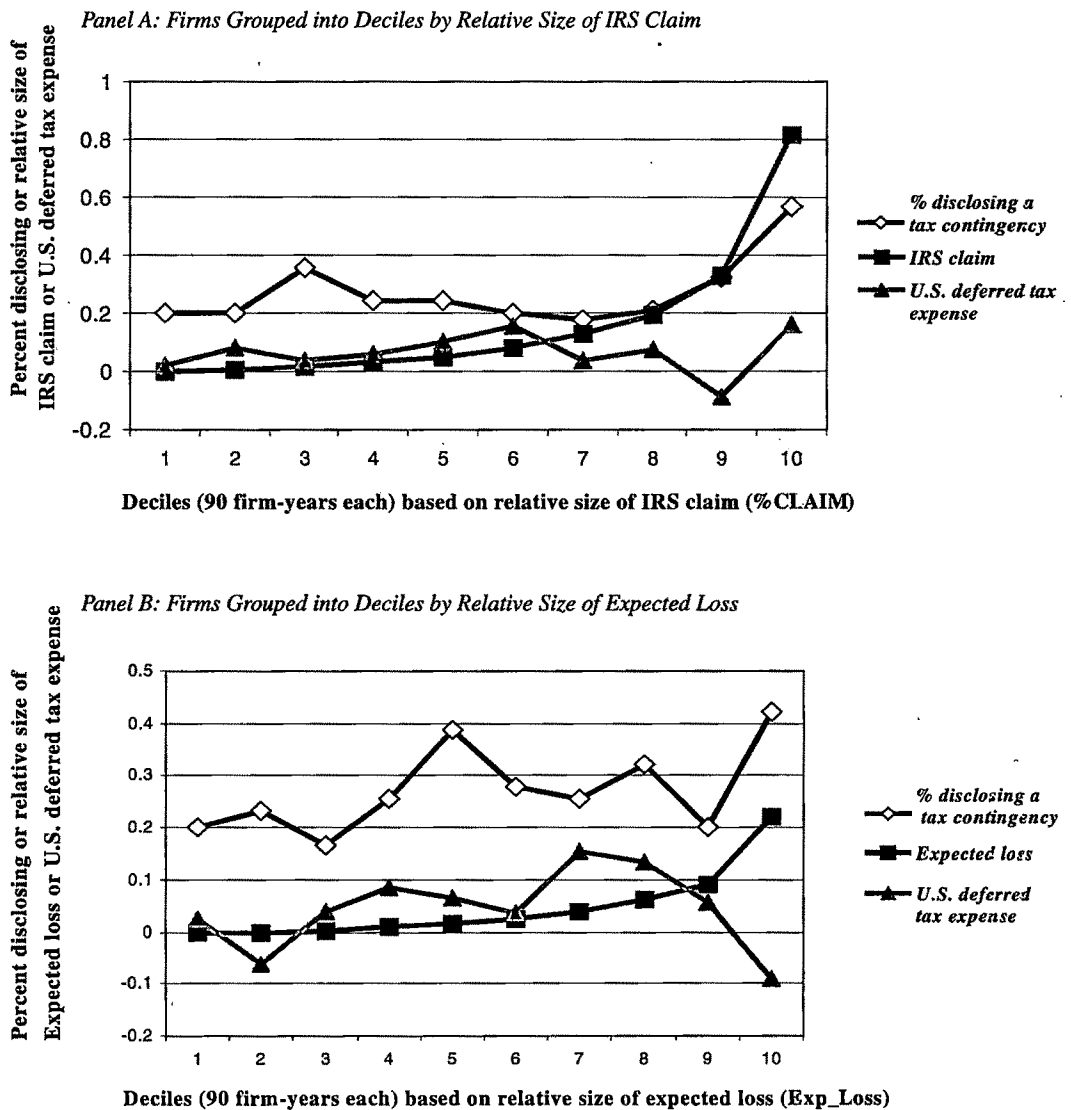
As expected, U.S. deferred tax expense is significantly smaller when firms DISCLOSE and is negatively correlated with CUSHION. Larger firms are somewhat more likely to disclose, but size is unrelated to CUSHION. Firms that issue new debt or equity are less likely to disclose contingent tax liabilities, while firms operating in litigious industries are more likely to do so. DISCLOS_QUAL is positively correlated with the amount of CUSHION (0.17, *p*-value = 0.000), consistent with these firms recording tax contingencies more conservatively.

Results: Disclosure of the Tax Contingency

Table 4 presents results from estimating the logistic regression (Equation [1]) of the disclosure of a tax contingency on the amount of the claim and our control variables. In Model A, we include the claim and control variables. In Model B, we use the expected loss rather than the claim as an explanatory variable. In Model C, we add DISCLOS_QUAL separately to Model A because this variable substantially restricts the sample size. Because disclosures can be sticky over time, and our measure of the claim accumulates over several years, we expect that the error terms are not independent across

¹⁰ Robustness tests that limit our sample to years (1991 to 1995) with similar accumulation periods yield identical inferences. The scaled claim levels out beginning in 1991 because the audits in our sample take approximately ten years to close. Thus, claims outstanding in 1991 begin accumulating from 1981. Although claims outstanding in 1995 include a few returns as early as 1981, the average accumulation would be from 1985.

FIGURE 1
Percent of Firms Disclosing Tax Contingencies, Median IRS Claim or Expected Loss and Median U.S. Deferred Tax Expense, Shown by Deciles of Claim (or Expected Loss)



See Table 1 for additional detail concerning variable definitions. The sample consists of nine years of data from 1987 to 1995 for the 100 largest industrial firms in the intersection of the Compustat PST file and the database of IRS audit and tax return data for the Coordinated Industry Cases (large-case audit firms).

TABLE 3

Means by Disclosure of Tax Contingencies (DISCLOSE), t-tests of Differences in Means, and Pearson Correlations between Accrual for Tax Contingency (CUSHION) with Unscaled and Scaled Claims (\$CLAIM, %CLAIM, and CLAIMASST), Expected Loss (EXP_LOSS), and Control Variables for the 100 Largest Firms in a Match of the Compustat PST File with the Coordinated Industry Cases Tax and Audit Data from 1987–1995

Variable ^a	n	Means Test			Correlation
		DISCLOSE = 1	DISCLOSE = 0	t-test of Differences	CUSHION
		Mean n = 655	Mean n = 245		Rho (p-value)
\$CLAIM	900	551.43	208.06	3.42***	-0.06 (0.09)
%CLAIM	900	0.35	0.13	5.42***	0.07 (0.05)
CLAIMASST	900	0.03	0.01	6.16***	0.06 (0.08)
EXP_LOSS	900	0.08	0.05	2.93***	0.22 (0.00)
USDEF_TAX	900	-0.08	0.09	-4.88***	-0.14 (0.00)
LOGASSETS	900	9.18	8.99	2.81***	-0.05 (0.15)
NEWCAP	900	0.29	0.40	3.04***	0.05 (0.12)
LITIGIOUS	900	0.23	0.07	5.45***	0.10 (0.00)
DISCLOS_QUAL	477	0.56	0.56	0.02	0.17 (0.00)

*** p-value less than 0.01.

^a See Table 1 for variable definitions.

observations for the same firm. We use Huber-White robust standard errors (Rogers [1993], generalizing White [1980]), which correct for dependence within a cluster of related observations and are robust to heteroskedasticity and serial correlation.¹¹

Model A shows that the probability that the firm discloses a contingent tax liability significantly increases in the amount of the claim ($z = 3.38$), consistent with H1. In Model B, the expected loss only weakly predicts disclosure ($z = 1.71$).

¹¹ We chose the Huber-White procedure over fixed-effects estimation because there is only modest variation in disclosure within firms. Inferences are unchanged if we estimate a fixed-effects model. Because fixed-effects estimation prevents us from exploring the effect of LITIGIOUS, which does not vary within-firm, we view this as a robustness test. The robust estimator does not assume that observations are independent. Instead, "assume, for a moment, that the observations denoted by j are not independent but that they can be divided into M groups $G_1, G_2, \dots, G_M$ that are independent. Then the robust estimator of variance is $\hat{V} = \hat{V}(\sum_{k=1}^M u_k^{(G)} u_k^{(G)}) \hat{V}$, where $u_k^{(G)}$ is the contribution of the k th group to the scores $\partial \ln L / \partial \beta$. That is, application of the robust variance formula merely involves using a different decomposition of, namely, $u_k^{(G)}$ $k = 1, \dots, M$ rather than u_j , $j = 1, \dots, N$ " (StataCorp 1999, [U] 23.11, p. 259).

TABLE 4
Pooled, Cross-Sectional Logistic Regressions of Disclosure of IRS Audit (DISCLOSE) on
Claim and Control Variables for the 100 Largest Firms in a Match of the Compustat PST
File with the Coordinated Industry Cases Tax and Audit Data from 1987-1995

$$\ln \frac{P_{disc}}{1 - P_{disc}} = \alpha + \beta X + \varepsilon \text{ where } P_{disc} = \frac{1}{1 + e^{-(\alpha + \beta X + \varepsilon)}}$$

= probability the firm will disclose a contingent tax liability.

$$\beta x = \beta_1 \%CLAIM_{i,t} \text{ (or } EXP\_LOSS_{i,t}) + \beta_2 ADEQUATE_{i,t} + \beta_3 USDEF\_TAX_{i,t} + \beta_4 LOGASSETS_{i,t} \\ + \beta_5 NEWCAP_{i,t} + \beta_6 LITIGIOUS_{i,t} + (\beta_7 DISCLOS\_QUAL_{i,t}) + \varepsilon_{i,t}$$

Variable ^b	Expected Sign	Model A	Model B	Model C
		Use %CLAIM as Explanatory Variable	USE EXP_LOSS as Explanatory Variable	Add DISCLOS_QUAL to Model A
		Coefficient (z-statistic) ^a	Coefficient (z-statistic) ^a	Coefficient (z-statistic) ^a
Intercept		-2.73 (-1.83)*	-3.44 (-2.08)**	-4.72 (-2.14)**
%CLAIM	+	1.96 (3.38)***		1.75 (2.34)**
EXP_LOSS	+		2.63 (1.71)*	
<i>Control Variables</i>				
ADEQUATE	-	-0.04 (-0.08)	-0.09 (-0.20)	0.11 (0.15)
USDEF_TAX	-	-0.77 (-2.22)**	-0.84 (-2.35)**	-1.84 (-3.25)***
LOGASSETS	+	0.14 (0.87)	0.26 (1.47)	0.35 (1.51)
NEWCAP	-	-0.39 (-1.37)	-0.55 (-1.90)*	-0.60 (-1.64)*
LITIGIOUS	+	1.56 (2.82)***	1.37 (2.52)**	1.49 (2.35)**
DISCLOS_QUAL	+			0.25 (0.34)
Observations		900	900	477
Pseudo R ²		12.9%	8.9%	18.6%
Wald Chi-Square		28.20	19.09	31.40

***, **, * p-value less than 0.01, 0.05, 0.10, respectively.

^a Huber-White robust standard errors correct for serial correlation among multiple-year observations per firm. See Rogers (1993) for a discussion of clustering dependent observations to correct standard errors.

^b See Table 1 for variable definitions.

Recording ADEQUATE cushion is unrelated to the probability of disclosure. As expected, USDEF_TAX is significantly negatively related to the probability of disclosure. This is consistent with the firm being less likely to disclose claims due to temporary differences than those due to permanent differences. Size is unrelated to the probability of disclosure in the logit model, but a firm is more likely to disclose if it operates in a litigious industry (LITIGIOUS), even after controlling for the amount of the claim.

Model C is estimated using approximately half our original sample, because FAF scores are not available for all firm-years of our sample. Results are similar to model A, except NEWCAP is now marginally significant. However, the coefficient on DISCLOS_QUAL is insignificantly different from 0.

We supplement our logistic regression of DISCLOSE on %CLAIM by using a piecewise linear function of %CLAIM. The piecewise linear regression estimates the change in the slope as we increase from one interval of %CLAIM to the next higher interval. We developed three separate piecewise linear regressions: one based on quintiles of the explanatory variable %CLAIM, one based on deciles of %CLAIM, and one based on 20 equal increments of 5 percentiles in %CLAIM. In untabulated results, the incremental coefficient is significant only in the highest group, in each of the three regressions. When we partition the data into quintiles, the z-statistic for the incremental effect of the top quintile of %CLAIM (which includes claims greater than 25 percent of normal income) is 3.61. When we partition the data into deciles, the z-statistic for the top decile of %CLAIM (which includes claims greater than 45 percent of normal income) is 2.79. Finally, when we partition the data into 20 equal increments, each representing 5 percentiles, the z-statistic for the top 5 percent of %CLAIM (which includes claims greater than 80 percent of normal income) is 2.89. These piecewise linear results show that the relation between disclosure and %CLAIM is limited to the largest claims, consistent with our previously reported descriptive evidence that the largest claims drive firms' decisions about disclosing contingent tax liabilities.

Results: Accrual of the Tax Cushion

Table 5 presents the results of estimating an OLS regression of the amount of the tax cushion on the expected loss and control variables. Similar to our logistic regressions, we begin with a basic model (Model A), then add a variable for DISCLOS_QUAL (Model B).

We find that the tax cushion is positively and significantly related to the expected loss, consistent with H2. This result is consistent with construct validity for both proxies (expected loss and tax cushion), but the relatively low R^2 is also consistent with measurement error. The tax cushion is about 88 percent of the expected loss in Model A, consistent with managers recording a lower estimate from a range rather than the midpoint of the expected loss. However, the cushion is roughly equal to the expected loss in Model B. When we use the firm's own average settlement percentage instead of the industry percentage to estimate expected loss, the results are qualitatively the same as in Table 5.

Model A shows that firms with high deferred taxes accrue less tax cushion, consistent with temporary differences generating fewer deficiencies that affect earnings. Firms with higher DISCLOS_QUAL (Model B) record more tax cushion, consistent with such firms being more conservative in recording accruals. In this subsample, firms that operate in litigious industries accrue more tax cushion.

Inferences Regarding Materiality

Using both the descriptive and multiple regression results, we now consider whether firms appear to satisfy the SFAS No. 5 requirements to disclose material tax contingencies.

TABLE 5
Pooled, Cross-Sectional Regressions of Estimated Accrual for Tax Contingency (CUSHION)
on Expected Losses and Control Variables for the 100 Largest Firms in a Match of the
Compustat PST File with the Coordinated Industry Cases Tax and Audit Data
from 1987–1995

$$Cushion_{i,t} = \beta_0 + \beta_1 EXP\_LOSS_{i,t} + \beta_2 USDEF\_TAX_{i,t} + \beta_3 LOGASSETS_{i,t} + \beta_4 NEWCAP_{i,t} \\ + \beta_5 LITIGIOUS_{i,t} + (\beta_6 DISCLOS\_QUAL_{i,t}) + \varepsilon_{i,t}$$

<i>Variable^a</i>	<i>Expected Sign</i>	<i>Model A Basic Model Coefficient (t-statistic)^a</i>	<i>Model B Add DISCLOS_QUAL Coefficient (t-statistic)^a</i>
Intercept		0.46 (1.28)	-0.33 (-0.82)
EXP_LOSS	+	0.88*** (2.51)	1.03** (2.10)
<i>Control Variables</i>			
USDEF_TAX	-	-0.14** (-1.94)	-0.02 (-0.15)
LOGASSETS	+	-0.02 (-0.51)	0.06 (1.30)
NEWCAP	-	0.05 (0.83)	-0.03 (-0.59)
LITIGIOUS	+	0.14 (1.22)	0.19* (1.62)
DISCLOS_QUAL	+		0.16** (1.97)
Observations		900	477
R ²		7.4%	16.4%

***, **, * p-value less than 0.01, 0.05, 0.10, respectively, one-tailed.

^a Huber-White robust standard errors are adjusted to correct for serial correlation among multiple-year observations per firm. See Rogers (1993) for a discussion of clustering dependent observations to correct standard errors.

^b See Table 1 for variable definitions.

Referencing Table 2, the mean %CLAIM for the no disclosure category (NONE) is about 13 percent. If the claims were all due to permanent differences that affected earnings, then this would suggest that firms are not disclosing all material claims. Using a conventional materiality threshold of 5 percent of normal income, if 40 percent or more of the tax deficiencies were due to permanent differences, then we would conclude that firms fail to disclose material tax contingencies.

If instead the claims are all due to temporary differences, then there is no earnings effect but only a reclassification from deferred to current tax liability. Thus, the claim relative to total assets is a more appropriate measure of materiality. The mean CLAIMASST

indicates that disclosures of DISPUTE occurred at conventional levels. The average disclosed dispute has a claim equal to 6 percent of assets. If the claims are entirely due to temporary differences, then the disclosures of MENTION and OPENYRS, with mean claims less than or equal to 2 percent of assets, do not appear necessary under conventional materiality thresholds.

Our logit regressions show that the probability of disclosure decreases in deferred taxes, consistent with a given level of claim being less material if it relates to temporary differences. Unfortunately, IRS data are not sufficiently fine to identify the portion of the claim that is due to permanent vs. temporary differences.

We use the estimated coefficients from the logistic regression to draw additional inferences. Using the medians for all variables and the coefficient estimates for Model A of Table 4, the predicted probability that the firm will disclose a contingent tax liability is 20 percent. If %CLAIM increases to 0.20 (the 3rd quartile) and all other variables remain at the median, then the predicted probability of disclosure increases only to 24 percent. The %CLAIM would have to increase to 0.77 to increase the predicted probability of disclosure to at least 50 percent. Combined with our piecewise linear regression results, we infer that, although the largest claims are more likely to trigger disclosure, many firms do not disclose material tax contingency claims.

Supplemental Tests

We next evaluate what scalar best reflects the relative materiality of the claim by comparing alternate scalars to the one used in our tables, the greater of pretax income or 5 percent of total assets.¹² First, we use pretax income as a scalar whenever it is positive, substituting 5 percent of total assets as a scalar only when the firm reports a loss for the year. We find that with this scalar, %CLAIM is no longer robustly related to the probability that the firm discloses a contingent tax liability. However, if we scale by the greater of pretax income or 10 percent of assets, or if we scale the claim by total assets, then our results are still significant and qualitatively unchanged from those reported in the tables. These results suggest that firms evaluate materiality relative to stable bases, ignoring downward fluctuations in income. Whether shareholders and policymakers would want the firm to disclose a claim that is material relative to reported income during a low-income year (but that is not material compared to normal income) is an interesting question for future research.

As previously discussed, because DISCLOSE and CUSHION may be endogenous, we conduct statistical tests of endogeneity. We perform a variant of the Hausman (1978) test, explained by Kennedy (1998).¹³ The Hausman test does not reject the null hypothesis that

¹² We also include the unscaled cumulative claim (\$CLAIM) as an additional explanatory variable in the estimation of Equation (1). The coefficient on \$CLAIM is insignificant and %CLAIM remains significant.

Our main cushion regressions scale all continuous variables by our normal income measure (the greater of pretax income or 5 percent of assets). As a supplemental test, we scale all variables by assets. Results are qualitatively similar to those in Table 5, with slightly stronger results on EXP_LOSS and LITIGIOUS. Finally, if we regress the unscaled cushion on unscaled explanatory variables, then the unscaled size effects dominate the regression and \$CLAIM is unrelated to CUSHION.

¹³ In this test we use the estimated coefficients from the regression models in Tables 4 and 5 to predict DISCLOSE and CUSHION in first-stage regressions. We include the predicted and actual values of CUSHION in the DISCLOSE logit regression. We include the predicted and actual values of DISCLOSE in the CUSHION regression. If DISCLOSE and CUSHION are endogenous, then the coefficients on the predicted values will be significant. The coefficients on predicted-DISCLOSE ($t = -0.33$) and predicted-CUSHION ($z = -0.79$) are insignificant, indicating no endogeneity.

the disclosure and accrual decisions are independent, consistent with our original specification and with interviews with audit and tax partners.

Because the claim (%CLAIM) and the expected loss (EXP_LOSS) are skewed, we re-estimate Equations (1) and (2) using the ranks of the explanatory variables and the continuous dependent variable (CUSHION). In Table 4, %CLAIM and EXP_LOSS become insignificant. This is consistent with Figure 1, which shows that the probability of disclosure increases only for the largest claims.¹⁴ In a ranked regression for Table 5, results become more significant, indicating that accruals are not as influenced by extreme expected losses.

Our main logistic regression classifies the dependent variable (DISCLOSE) equal to 1 if the firm makes any disclosure of contingent tax liabilities. We re-estimate the logistic regression classifying the dependent variable equal to 1 only when firms disclose a DISPUTE.¹⁵ %CLAIM and LITIGIOUS remain equally significantly related to disclosure. However, USDEF_TAX is unrelated to disclosing a dispute.

V. CONCLUSIONS

We examine the relation between materiality and contingent liability accounting. Because data on outside parties' contingent claims against firm assets are not generally available, there is little archival empirical evidence on the level of materiality that prompts firms to disclose contingent liabilities. We use confidential IRS data on tax deficiencies proposed during tax examination to measure contingent tax liabilities.

We find that firms infrequently disclose the existence of IRS claims, and that the majority of disclosures do not contain the detailed information required by SFAS No. 5. Nevertheless, we find that the probability of disclosure increases with the amount of the claim. It is more difficult to identify the threshold at which firms disclose claims, although piecewise regression results indicate that only the largest claims trigger disclosure. Using a conventional 5-percent-of-earnings or -assets materiality threshold, our descriptive evidence is consistent with firms disclosing too little if the contingency affects only earnings, and firms disclosing more than expected if the contingency affects only balance sheet classification. Because prior research indicates that contingent tax liabilities generate both earnings and balance sheet effects, either extreme is unrealistic. However, unless the majority of IRS claims are due to temporary differences, firms appear to underdisclose material claims, even if the basis for materiality judgments is the amount of the expected loss rather than the amount of the claim.

Our evidence further suggests that firms determine materiality with respect to stable measures rather than only current-year income. We believe future research should investigate whether an item that is immaterial compared to "normal income" (the greater of income or 5 percent of assets) but material compared to current (low) income would affect financial statement users' decisions.

Although SFAS No. 5 mandates disclosure and recording of loss contingencies, we learn that discretionary disclosure variables affect these decisions. In particular, we find that firms in litigious industries are more likely to disclose material tax contingencies. For

¹⁴ We also estimated the logistic regression using LOGCLAIM, defined as the natural logarithm of $(1 + \%CLAIM)$, in place of %CLAIM. The coefficient on LOGCLAIM is significant at the 0.01 level in the main regression ($z = 3.26$) and at the 0.05 level in the subsample with the disclosure quality data ($z = 2.20$).

¹⁵ We also estimate unordered and ordered multinomial logit regressions. We believe that the binary DISCLOSE dependent variable best addresses our hypothesis about disclosure under SFAS No. 5. However, the multinomial regressions yield expected results. DISPUTE is the only separate disclosure category significantly related to %CLAIM, and firms in LITIGIOUS industries more frequently disclose disputes. USDEF_TAX, NEWCAP, and LITIGIOUS predict OPENYRS in the expected direction. The regression for MENTION is largely uninformative, but this category has a small sample size.

those firm-years for which we have disclosure-quality rankings by the FAF, we find that firms issuing new capital are less likely to disclose tax contingencies and that firms with high FAF rankings are more likely to accrue tax contingencies.

Finally, as expected, we find that the tax cushion, computed as the difference between current tax expense and tax on the return, increases with the amount of the expected loss. This is consistent with the requirements under SFAS No. 5 that firms must accrue a liability for material probable and estimable losses.

APPENDIX

Examples below are for firms that are *not* in our sample (due to confidentiality), but are characteristic of disclosures made by our sample firms.

MENTION:

Wabash National Corporation, December 31, 1998 Annual Report

Various lawsuits, claims, and proceedings have been or may be instituted or asserted against the Company arising in the ordinary course of business, including those pertaining to product liability, labor- and health-related matters, successor liability, and possible tax assessments. None of these claims are expected to have a material adverse effect on the Company's financial position or its annual results of operations.

OPENYRS:

Magellan Health Services Inc., September 30, 1998 Annual Report

The Internal Revenue Service is currently examining the Company's income tax returns for fiscal 1992 and 1993 and expects a report to be issued during fiscal 1999. In management's opinion, adequate provisions have been made for any adjustments, which may result from these examinations, including a potential reduction in the amount of NOL carryforwards. The Company believes the examinations could result in a reduction in NOL carryforwards available to offset future taxable income.

Morrison Knudsen Corporation, November 30, 1998 Annual Report

Years prior to 1994 are closed to examination for federal tax purposes. Management believes that adequate provision has been made for probable tax assessments.

DISPUTE:

Unocal Corporation, December 31, 1997 Annual Report

In December 1994, the company received a Notice of Proposed Deficiency (Notice) from the Internal Revenue Service (IRS) related to the years 1985 through 1987. In February 1995, the company filed a protest of the proposed tax deficiency with the Appeals section of the IRS. Discussions with the Appeals Officer are nearly complete, and it now appears unlikely that any issues raised in the Notice will proceed to either litigation or mediation, and it is expected that all matters will be settled. The settlement will require approval by the Joint Committee on Taxation of the U.S. Congress and such approval should be granted.

Amdahl Corporation, December 29, 1995 Annual Report

In the fourth quarter of 1994 the IRS issued a notice of deficiency to the Company for disputed items related to the 1983 through 1986 tax years, the most significant of which related to the treatment of system spares. The proposed tax deficiency totals approximately 40,200,000 and would carry interest through December 29, 1995 of approximately

78,000,000. If paid, the tax and interest would give rise to a deferred tax asset of approximately 51,000,000, subject to the recognition criteria of FAS No. 109. State income taxes payable as a result of the proposed tax deficiency would be approximately 15,700,000, net of federal income tax benefit.

In the first quarter of 1995 the Company filed a petition in the United States Tax Court contesting the proposed deficiency. Management believes the Company possesses strong factual support for its treatment of system spares and will vigorously defend its position. In the opinion of management, the final resolution of the proposed deficiency will not have a material adverse impact on the Company's financial position or results of operations. The IRS field audit of the Company's 1987 through 1990 tax years is in progress.

REFERENCES

- Ali, A., and S. Kallapur. 2001. Securities price consequences of the Private Securities Litigation Reform Act of 1995 and related events. *The Accounting Review* (July): 431–460.
- American Institute of Certified Public Accountants (AICPA). 1996. *Environmental Remediation Liabilities*. Statement of Position No. 96-1. New York, NY: AICPA.
- Banks, D. W., and W. R. Kinney, Jr. 1982. Loss contingency reports and stock prices: An empirical study. *Journal of Accounting Research* (Spring): 240–254.
- Barth, M. E., and M. F. McNichols. 1994. Estimation and market valuation of environmental liabilities relating to Superfund sites. *Journal of Accounting Research* (Supplement): 177–209.
- , ———, and G. P. Wilson. 1997. Factors influencing firms' disclosure about environmental liabilities. *Review of Accounting Studies* 2: 35–64.
- Bates, H. L., R. W. Ingram, and P. M. J. Reckers. 1982. Auditor-client affiliation: The impact on materiality. *Journal of Accountancy* (April): 60–63.
- Botosan, C. A. 1997. Disclosure level and the cost of equity capital. *The Accounting Review* (July): 323–350.
- , and M. Plumlee. 2001. A re-examination of disclosure level and the expected cost of equity capital. Working paper, University of Utah, Salt Lake City, UT.
- Chewning, G., K. Pany, and S. Wheeler. 1989. Auditor reporting decisions involving accounting principle changes: Some evidence on materiality thresholds. *Journal of Accounting Research* (Spring): 78–96.
- Cloyd, B. 1995. The effects of financial accounting conformity on recommendations of tax preparers. *The Journal of the American Taxation Association* 17 (Fall): 50–70.
- , J. Pratt, and T. Stock. 1996. The use of financial accounting choice to support aggressive tax positions: Public and private firms. *Journal of Accounting Research* 34 (Spring): 23–43.
- Dworin, L. 1985. On estimating corporate tax liabilities from financial statements. *Tax Notes* (December 2): 965–971.
- Financial Accounting Standards Board (FASB). 1975. *Accounting For Contingencies*. Statement of Financial Accounting Standards No. 5. Norwalk, CT: FASB.
- . 1976. *Reasonable Estimation of the Amount of a Loss: An Interpretation of FASB Statement No. 5*. FASB Interpretation No. 14. Norwalk, CT: FASB.
- . 1992. *Accounting for Income Taxes*. Statement of Financial Accounting Standards No. 109. Norwalk, CT: FASB.
- Francis, J., D. Philbrick, and K. Schipper. 1998. Earnings surprise and litigation risk. *The Journal of Financial Statement Analysis* 3: 15–27.
- Frankel, R., M. McNichols, and G. P. Wilson. 1995. Discretionary disclosure and external financing. *The Accounting Review* (January): 135–150.
- Frost, C. A. 1991. Loss contingency reports and stock prices: A replication and extension of Banks and Kinney. *Journal of Accounting Research* (Spring): 157–169.
- Hanlon, M., and T. Shevlin. 2001. Accounting for tax benefits of employee stock options and implications for research. Working paper, University of Washington, Seattle, WA.
- Hausman, J. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271.

- Internal Revenue Service (IRS). 1995. Corporate voluntary compliance baseline measures. Task Force Draft Report, Coordinated Examination Program (October). Washington, D.C.: Government Printing Office.
- Kennedy, P. 1998. *A Guide to Econometrics*. Cambridge, MA: MIT Press.
- Kieso, D. E., and J. J. Weygandt. 1998. *Intermediate Accounting*. 9th edition. New York, NY: John Wiley & Sons, Inc.
- Lang, M., and R. Lundholm. 1993. Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of Accounting Research* (Autumn): 246–271.
- , and ———. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* (October): 467–492.
- Leslie, D. A., A. D. Teitlebaum, and R. J. Anderson. 1980. *Dollar-Unit Sampling: A Practical Guide for Auditors*. Toronto, Canada: Pitman Publishing Ltd.
- . 1985. *Materiality: The Concept and Its Application to Auditing*. Toronto, Canada: The Canadian Institute of Chartered Accountants.
- Levitt, A. 1998. The numbers game. Remarks delivered at the NYU Center for Law and Business, New York, NY, September 28.
- Libby, R., and W. Kinney. 2000. Does mandated audit communication reduce opportunistic corrections to manage earnings to forecasts? *The Accounting Review* (October): 383–404.
- Li, Y., G. Richardson, and D. Thornton. 1997. Corporate disclosure of environmental liability information: Theory and evidence. *Contemporary Accounting Research* 14 (3): 435–474.
- Liu, C., S. G. Ryan, and J. M. Wahlen. 1997. Differential valuation implications of loan loss provisions across banks and fiscal quarters. *The Accounting Review* (January): 133–146.
- Miller, G. S., and J. D. Piotroski. 2000. Forward-looking earnings statements: Determinants and market response. Working paper, University of Chicago, Chicago, IL.
- Mills, L. F. 1998. Book-tax differences and Internal Revenue Service adjustments. *Journal of Accounting Research* (Autumn): 343–356.
- , and K. J. Newberry. 2001. The influence of tax and nontax costs on book-tax reporting differences: Public and private firms. *The Journal of the American Taxation Association* (Spring): 1–19.
- Moriarty, S., and F. Barron. 1976. Modeling the materiality judgments of audit partners. *Journal of Accounting Research* (Autumn): 611–618.
- Nelson, M. W., J. Elliott, and R. Tarpley. 2001. Evidence from auditors about managers' and auditors' earnings-management decisions. Working paper, Cornell University, Ithaca, NY.
- Patillo, J. W. 1976. *The Concept of Materiality in Financial Reporting*. New York, NY: Financial Executives Research Foundation.
- Petroni, K. R. 1992. Optimistic reporting in the property-casualty insurance industry. *Journal of Accounting & Economics* (December): 485–508.
- Rogers, W. 1993. Regression standard errors in clustered samples. In *Stata Technical Bulletin Reprints*, Vol. 13. College Station, TX: Stata Press.
- Securities and Exchange Commission (SEC). 1993. *Environmental Liabilities*. SEC Staff Accounting Bulletin No. 92. Washington, D.C.: SEC.
- . 1999. *Materiality*. SEC Staff Accounting Bulletin No. 99. Washington, D.C.: SEC.
- Skinner, D. J. 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research* (Spring): 38–60.
- . 1997. Earnings disclosures and stockholder lawsuits. *Journal of Accounting & Economics* (November): 249–282.
- Smith, S. 2001. Corporate tax aggressiveness and booking a tax cushion: An experiment. Working paper, Western Washington University, Bellingham, WA.
- StataCorp. 1999. *Stata Statistical Software: Release 6.0. User's Guide*. [U] 23.11: 256–260. College Station, TX: Stata Press.
- Turner, L. 2000. A QT report card for high quality financial reporting. Remarks at the Hylton Lecture Series in Accountancy: Critical Issues in Accounting Forum, Wake Forest University, Winston-Salem, NC, April 25.

- Wahlen, J. M. 1994. The nature of information in commercial bank loan loss disclosures. *The Accounting Review* (July): 455–478.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* (May): 817–838.

The Effect of Mandated Market Risk Disclosures on Trading Volume Sensitivity to Interest Rate, Exchange Rate, and Commodity Price Movements

Thomas J. Linsmeier
Michigan State University

Daniel B. Thornton
Queen's University

Mohan Venkatachalam
Stanford University

Michael Welker
Queen's University

ABSTRACT: We hypothesize that firms' 10-K market risk disclosures, recently mandated by SEC Financial Reporting Release No. 48 (FRR No. 48), reduce investors' uncertainty and diversity of opinion about the implications, for firm value, of changes in interest rates, foreign currency exchange rates, and commodity prices. We argue that this reduced uncertainty and diversity of opinion should dampen trading volume sensitivity to changes in these underlying market rates or prices. Consistent with this hypothesis, we find that after firms disclose FRR No. 48-mandated information about their exposures to interest rates, foreign currency exchange rates, and energy prices, trading volume sensitivity to changes in these underlying market rates and prices declines, even

The authors acknowledge funding from the Social Sciences and Humanities Research Council of Canada and Queen's University School of Business. We appreciate the helpful comments of Anwer Ahmed, Linda Bamber, Orie Barron, Robert Bayless, Steve Fortin, Neil Pearson, Linda Vincent, two anonymous reviewers, and workshop participants at the University of Alberta, Concordia University, Syracuse University, and the Joint Symposium of the 11th Annual Financial Economics and Accounting Conference and the 7th Mitsui Life Symposium on Global Financial Markets (held at the University of Michigan). Paul Buchanan, Eric Chang, and Lon Draper provided able research assistance.

Submitted July 2000
Accepted August 2001

after controlling for other factors associated with trading volume. The observed declines in trading volume sensitivity are consistent with FRR No. 48 market risk disclosures providing useful information to investors.

Keywords: *disclosure regulation; market risk disclosures; derivatives; trading volume; investors' uncertainty; diversity of opinion.*

Data Availability: *The data used in this study are publicly available from the sources indicated in the study. A list of sample firms is available from the authors.*

I. INTRODUCTION

The Securities and Exchange Commission (SEC) recently issued Financial Reporting Release No. 48 (SEC 1997, hereafter "FRR No. 48"), mandating forward-looking, quantitative market risk disclosures in companies' 10-K reports. FRR No. 48 defines market risk as the risk of loss due to adverse changes in market rates or prices, such as interest rates, foreign currency exchange rates, and commodity prices. Following highly publicized derivative-related losses in the 1990s, FRR No. 48 responds to appeals from investors and other constituents for enhanced public disclosure of firms' exposures to market risk (Linsmeier and Pearson 1997). This study provides evidence relating to the SEC's (1997) claim that FRR No. 48 disclosures provide investors with useful information.

We posit that FRR No. 48 disclosures can be useful to investors in at least two ways. First, at the release of the 10-K report, investors can use the information contained in the disclosures to revise or confirm their expectations about a firm's exposures to interest rate, foreign currency exchange rate, and/or commodity price changes. (Hereafter we refer to such changes collectively as "underlying market rate/price changes.") To the extent that disclosure alters market expectations, it can lead to a one-time change in the firm's stock price, systematic risk, or stock price sensitivity to underlying market rate/price changes. Second, investors can use their revised estimates of risk exposures to continuously assess the implications for firm value of underlying market rate/price changes that occur anytime after the 10-K release date. To the extent that this second, continuous use of the information contained in FRR No. 48 disclosures reduces investors' uncertainty and diversity of opinion about the implications of underlying market rate/price changes for firm value, Kim and Verrecchia's (1994) theoretical model suggests that we should observe less trading volume sensitivity to underlying market rate/price changes after the 10-K release date. Our analysis focuses on this second disclosure use.

As predicted, for a broad sample of nonfinancial firms, we document a decline in trading volume sensitivity to underlying market rate/price changes after disclosure of FRR No. 48 market risk information. We do *not* find a similar decrease in trading volume sensitivity after 10-K filings by two control samples: (1) the same firms in the prior fiscal year, before FRR No. 48 disclosures were required, and (2) a sample of different, nondisclosing firms in the same fiscal year in which the disclosing firms made their first FRR No. 48 disclosures. These results are consistent with the hypothesis that FRR No. 48 disclosures provide investors with useful information, which reduces their uncertainty and diversity of opinion about the implications of underlying market rate/price changes for firm value.

The paper continues as follows. The next section describes FRR No. 48 disclosures and cites *a priori* arguments for and against their usefulness. The remaining sections review related research, develop hypotheses, describe the research design, outline sample selection, and present empirical results. The paper concludes with an interpretation of results, a summary of limitations, and suggestions for future research.

II. FRR NO. 48 BACKGROUND AND DESCRIPTION

The introduction to FRR No. 48 states that the disclosures “are designed to provide additional information about market risk sensitive instruments, which investors can use to better understand and evaluate the market risk exposures of a registrant” (SEC 1997, 3). The SEC is concerned with providing full and fair disclosure to individual investors and ensuring that they “are on a level playing field with market insiders” (SEC 2000, 2) because “today’s self-directed, online investors do not expect to rely exclusively on research and analysis performed by professionals” (SEC 2000, 3). In FRR No. 48, the SEC pledged to monitor the effectiveness of the disclosure requirements and to amend them if appropriate. It also promised Congress that it would conduct a review of the effects of market risk disclosures after three years had passed. While that review progresses, existing academic literature, reviewed below, is inconclusive about whether market risk disclosures are useful to investors.

Financial economics literature suggests that market risk disclosures can reduce information asymmetries between managers and investors, if firms manage risk for reasons that are not transparent to investors. In perfect capital markets, firm-level hedging of interest rates, foreign currency exchange rates, or commodity prices is inefficient because investors can better diversify these risks on personal account, according to their own risk preferences (DeMarzo and Duffie 1991). In imperfect markets, however, firm-level risk management can enhance firm value in several ways. These include reducing anticipated tax payments, reducing the likely costs of financial distress, protecting investment, responding to managerial risk aversion, and increasing perceived firm value by preserving a trend of smoothly increasing earnings (Smith and Stultz 1985; Nance et al. 1993; DeMarzo and Duffie 1995; Mian 1996; Tufano 1996; Barth et al. 1999). If investors do not directly observe firms’ risk management activities, then quantitative, forward-looking disclosure of the exposures resulting from these activities has the potential to convey useful information to the market and enhance investors’ understanding of firms’ market risk exposures.

FRR No. 48 requires firms to disclose quantitative market risk information in their 10-K filings for each material category of market risk (e.g., interest rate risk, foreign currency exchange rate risk, commodity price risk). Within each risk category, FRR No. 48 allows companies to present quantitative market risk information using three alternative formats:

- 1) *Tabular presentation* of fair values and contract terms sufficient to determine market risk-sensitive instruments’ future cash flow amounts by expected maturity dates.
- 2) *Sensitivity analyses* describing the possible effect on earnings, cash flows, or fair values from selected, hypothetical changes in underlying market rates/prices.
- 3) *Value-at-risk* (VAR) disclosures expressing the potential loss in earnings, cash flows, or fair values from underlying market rate/price changes with a selected probability of occurring.

Firms may use different quantitative disclosure formats across exposure categories.¹ The Appendix provides an example of each disclosure format.

Consistent with the goal of mandating disclosures relating to all market risk-sensitive instruments, FRR No. 48 requires 10-K disclosure of market risks inherent in derivatives and in all other, nonderivative financial instruments included within the scope of SFAS No. 107 (FASB 1991a), e.g., investments, loans, structured notes, and debt obligations. FRR No. 48 (SEC 1997, Section II) reminds SEC registrants that market risk exposures may

¹ FRR No. 48 also requires separate reporting for trading and nontrading instruments; however, this study focuses on nonfinancial firms, which generally do not hold significant amounts of trading instruments.

arise from instruments and positions that are outside the explicit scope of the rule, such as commodity positions and cash flows from anticipated transactions. It encourages companies either to include such instruments and positions in their disclosures voluntarily or to discuss their exclusion as a disclosure limitation.

The SEC required two initial groups of firms, thought to be most familiar with measuring and managing market risk, to comply with FRR No. 48 in fiscal years ending after June 15, 1997: (1) bank and thrift institutions, and (2) nonfinancial enterprises with equity-market capitalization exceeding \$2.5 billion on January 28, 1997. All other firms must comply with FRR No. 48 in SEC filings for fiscal years ending after June 15, 1998. We focus on the first 10-K disclosures of nonfinancial firms required to comply with FRR No. 48 in fiscal years ending after June 15, 1997.²

We evaluate the incremental effect of FRR No. 48 in light of the disclosure requirements that preceded it. By 1997, accounting standard setters had been considering firms' financial reporting for financial instruments for nearly a decade.³ Several highly publicized derivative-related losses in 1994 accelerated the pace of accounting pronouncements (Baliga 1994; Loomis 1995). Yet, even after the publication of SFAS No. 119 (FASB 1994d),⁴ a pronouncement intended to enhance firms' disclosures about derivatives, an internal review of approximately 500 10-K filings (SEC 1995) found two deficiencies in market risk disclosures:

- 1) *Qualitative deficiencies*: Prior to FRR No. 48, the only mandated, forward-looking market risk disclosures were qualitative, and they pertained only to derivative financial instruments issued for purposes other than trading. For these instruments, SFAS No. 119 required disclosure of "the entity's objectives for holding or issuing the derivative financial instruments, the context needed to understand those objectives, and its general strategies for achieving those objectives" (FASB 1994d, para. 11a). However, by focusing on certain derivatives in isolation, these disclosures did not give investors a complete understanding of a firm's risk management objectives and strategies. FRR No. 48 addresses this omission by requiring disclosure of similar qualitative information for a broader set of instruments, including nonderivative financial instruments and derivative commodity instruments.
- 2) *Quantitative deficiencies*: Prior to FRR No. 48, quantitative disclosures relating to market risk-sensitive instruments were very limited. SFAS No. 119 mandated disclosure of notional amounts of derivative financial instruments (FASB 1994d, para. 8b), while Rule 5-02 of Regulation S-X required disclosure of the nature and terms of debt obligations. The SEC noted, however, that such information:

(i) often is abbreviated, (ii) is presented piecemeal in different parts of the financial statements, and (iii) does not apply to all market-risk-sensitive instruments. Thus, investors often are unable to assess whether or how particular financial and commodity instruments affect a registrant's net market risk exposure. (SEC 1995, section IIIB1)

² We exclude financial service institutions because regulators required them to make similar disclosures before FRR No. 48 and prior studies have already examined them (e.g., Barth et al. 1996; McNally 1996; Collins and Venkatachalam 1996; Venkatachalam 1996; Schrand 1997).

³ See, for example, FASB (1990, 1991a, 1991b, 1991c, 1994a, 1994b, 1994c, 1994d, 1995, 1996a, 1996b, 1996c, 1997). Later standards (FASB 1998, 1999, 2000) refined the accounting for derivatives and hedging activities but did not expand disclosure requirements for prospective information regarding firms' market risk exposures.

⁴ SFAS No. 133 (FASB 1998) and SFAS No. 138 (FASB 2000) superseded SFAS No. 119.

Prior to FRR No. 48, then, investors had insufficient information to assess firms' exposures to market risks.⁵ Quantitative information about the market risks inherent in non-derivative positions was very limited, making it difficult to assess how derivatives altered firms' net exposures to underlying market rate/price changes. Reporting on a survey of investors and creditors, the AICPA's Special Committee on Financial Reporting stated, "[u]sers...complain that business reporting is not answering important questions, such as: What [innovative financial] instruments has the company entered into, and what are their terms?...What risks has the company transferred or taken on?" (AICPA 1994, 76). FRR No. 48 seeks to redress these shortcomings by mandating the first forward-looking, quantitative disclosures about firms' net exposures to market risks.

In contrast, critics of FRR No. 48 cite two main reasons that the disclosures are unlikely to be informative. First, managers can choose the format of quantitative disclosures and can base their disclosures on questionable assumptions about future events and actions. Even a January 1997 internal memo from the SEC chief economist suggests that subjectivity in quantitative risk measurement allows managers to misrepresent companies' net market risk exposures, potentially misleading investors (Beckett 1997; Logan and Montgomery 1997). Second, investors are likely to have difficulty processing the quantitative, probabilistic information contained in some FRR No. 48 disclosures and comparing the exposures of companies making tabular, sensitivity, and VAR disclosures (Hodder et al. 2001). Therefore, critics of FRR No. 48 argue that the disclosures may not improve investors' assessments of firms' market risk exposures.

In July 1998, the SEC staff reported to Congress on first-year FRR No. 48 filings, including the results of a survey of a limited number of analysts and investors concerning their experiences with FRR No. 48. Responses suggest that the disclosures provide new and useful information. The SEC staff report stated, "Investors anticipate that in the future they will compare a registrant's market risk disclosures to actual results, the registrant's past disclosures and the disclosures of other registrants" (SEC 1998, 9–10). However, survey respondents also suggested that it was "too early to tell whether the disclosures are resulting in more efficient markets" (SEC 1998, 2).

We provide evidence relating to the divergent claims of FRR No. 48's supporters and critics by developing and testing hypotheses about the effect of FRR No. 48 disclosures on the sensitivity of trading volume to underlying market rate/price changes. Before developing the hypotheses, we briefly review related market risk research.

III. MARKET RISK RESEARCH

Two streams of research on market risk exposures are germane to this study. Studies in the first stream establish that stock prices are sensitive to changes in interest rates, foreign currency exchange rates, and commodity prices. They also assess the degree of cross-sectional association between stock price sensitivities to these underlying market rate/price changes and limited amounts of market risk information available to researchers prior to FRR No. 48. The second stream specifically examines the usefulness of information mandated by FRR No. 48.

⁵ SFAS No. 119 encouraged, but did not require, disclosure of quantitative information about market risk inherent in derivative financial instruments and other financial instruments. According to the SEC, prior to FRR No. 48, firms virtually never disclosed such information voluntarily (SEC 1995, Section IIB1a).

Five studies in the first stream are pertinent here. Schrand (1997), using private, regulatory data, shows that savings and loan institutions' use of interest rate derivatives dampens their equity price sensitivity to interest rate changes. Tufano (1998), using analyst-generated private information, shows that commodity price hedging lowers gold mines' stock price sensitivity to gold-price changes. Jorion (1990), using public disclosures, finds that firms' stock price sensitivity to the U.S. dollar price of a trade-weighted basket of foreign currencies is greater for firms with more foreign operations. Bartov and Bodnar (1994) document that the market does not fully incorporate information about firms' sensitivity to foreign currency exchange rate fluctuations until that sensitivity affects reported accounting earnings. Their results suggest that extant public disclosures of the extent of foreign operations provided only incomplete measures of firms' sensitivities to currency risk. Finally, Wong (2000) documents weak and inconsistent evidence of an association between firms' stock price sensitivity to foreign currency exchange rate changes and their derivative positions reported according to the requirements of SFAS No. 119 (FASB 1994d).

Studies in this first research stream show that security returns are sensitive to changes in underlying market rates/prices. They also show that, prior to FRR No. 48, firms' stock price sensitivities to changes in underlying market rates/prices were associated with limited amounts of mostly private, industry-specific market risk information. However, consistent with the SEC review of public disclosures, they find that SFAS No. 119 information was not very useful to investors. None of these studies focus on the usefulness of FRR No. 48 information. However, Schrand (1997), who uses regulatory data similar to FRR No. 48 tabular disclosures, suggests that FRR No. 48 disclosures may be useful to investors. Similarly, Wong (2000) suggests that the more comprehensive disclosures required by FRR No. 48 would better portray firms' market risk exposures and overcome some of the deficiencies that he identified in SFAS No. 119 information.

The second stream of related literature includes four studies explicitly examining the usefulness of FRR No. 48 information. However, the first three do not employ actual FRR No. 48 information. Instead, they develop proxies for FRR No. 48 disclosures from information that oil and gas firms and banks disclosed before FRR No. 48 became effective. For a sample of oil and gas producers, Rajgopal (1999) constructs proxies for FRR No. 48's tabular and sensitivity disclosures from disclosures of reserve fair values (SFAS No. 69, FASB 1982) and derivative notional amounts (SFAS No. 119, FASB 1994d). He reports positive associations between FRR No. 48 disclosure proxies and firms' stock return sensitivities to oil and gas price changes. Rajgopal and Venkatachalam (2000), for a sample of 25 U.S. petroleum refiners, assess the cross-sectional relation between earnings sensitivities and returns sensitivities to oil prices. They find a positive relation between a proxy for FRR No. 48 earnings sensitivity disclosures and market perceptions of risk exposure. Ahmed et al. (2000) examine commercial banks' maturity gap disclosures, filed with regulators in call reports, as proxies for FRR No. 48 tabular disclosures. They find that these disclosures are useful in predicting changes in banks' net interest income. Finally, Thornton and Welker (2000) examine the effect of actual FRR No. 48 disclosures on market assessments of firms' equity price sensitivities to oil and gas price changes. They find that oil and gas producers disclosing sensitivity analyses or VAR experience greater post-FRR No. 48 changes in their equity price sensitivity to oil and gas price changes than do tabular disclosers and nondisclosers.

In sum, this second research stream suggests that FRR No. 48 market risk disclosures provide useful information to investors. However, the evidence to date is based mostly on proxies for FRR No. 48 information, and is limited to firms in the oil-and-gas and banking industries. In contrast, we examine the market effects of actual FRR No. 48 disclosures,

using a broad sample of nonfinancial firms. Moreover, existing research primarily examines the relation between *stock price* sensitivity to underlying market rate/price changes and market risk exposure information. In contrast, we examine the sensitivity of *daily trading volume* to underlying market rate/price changes. In the next section, we explain why we expect this volume sensitivity to be positively related to investors' uncertainty and diversity of opinion about firms' market risk exposures, and why we predict that this trading volume sensitivity declines after FRR No. 48 disclosures.

IV. HYPOTHESIS DEVELOPMENT

Prior research concludes that trading volume increases with investor uncertainty and diversity of opinion (Karpoff 1986; Kim and Verrecchia 1994; Kandel and Pearson 1995; Barron 1995; Bamber et al. 1997, 1999). Kim and Verrecchia's (1994) model of market liquidity and trading volume provides a conceptual foundation for our hypothesis development.

Kim and Verrecchia (1994)

Kim and Verrecchia's (1994) model captures two salient features of the disclosure setting: uncertainty (defined as imprecision of public information) and diversity of opinion (defined as 1 minus the correlation between the information that any two information processors obtain). In the model, market participants receive a signal containing imprecise information about a firm's uncertain future cash flows, and decide whether to incur costs in further processing the value-relevant information in the signal. Although Kim and Verrecchia (1994) (hereafter KV) apply their model to the earnings release signal containing information about the firm's anticipated cash flows, we apply the model to the change in underlying market rates/prices signal (such as a change in interest rates, foreign currency exchange rates, or commodity prices). Movements in underlying market rates/prices *per se* are not firm-specific signals; however, prior literature suggests that they have firm-specific implications because they are associated with equity returns and earnings (Jorion 1990; Schrand 1997; Tufano 1998; Ahmed et al. 2000). We posit that the firm-specific implications of these movements for anticipated future cash flows were unclear prior to FRR No. 48, when investors had little or no information about firms' net market risk exposures.

The signal in the KV model stimulates investors to engage in costly information processing because the implications of the signal for the firm's anticipated cash flows are subject to interpretation. Investors therefore perceive that they can obtain an informational advantage through costly information processing. The number of investors choosing to incur costs to process information, in an attempt to profit through informed trade, increases with the informational advantage that they perceive they can obtain from processing information. Less precise public information increases the perceived returns to determining the firm-specific implications of a signal, increasing the number of investors willing to process the information and trade. The potential to profit from informed trade, and hence the likely benefit of costly information processing, also increases with the extent to which market participants disagree (i.e., have diversity of opinion) about the valuation implications of the signal.

Hypotheses

We posit that when investors learn of an underlying market rate/price change, such as a change in interest rates, they decide whether it is cost effective to seek an informational advantage by engaging in costly information processing to better determine the firm-specific cash flow implications of the rate/price change. The SEC's initial review of the effects of

FRR No. 48 disclosures, described in Section II, suggests that before firms made the disclosures, there was a dearth of information about individual firms' net market risk exposures. Therefore, (1) the firm-specific cash flow implications of underlying market rate/price changes tended to be imprecise, and (2) because of this imprecision, investors who engaged in costly, private processing of the underlying market rate/price change information were likely to have diverse opinions about its implications for firm value. For example, prior to FRR No. 48 disclosures, investors seeing a ten-basis-point increase in interest rates would tend to have imprecise assessments of and diverse opinions about firm-specific implications of the change.

Smithson et al. (1995) argue that changes in interest rates, foreign currency exchange rates, and commodity prices generally follow a random-walk process. This suggests that, generally, investors view the most recent changes in underlying market rates/prices as permanent ones affecting the magnitude of a firm's expected future cash flows. The potential effect on firm value of an underlying market rate/price change, therefore, should increase with the magnitude of the rate/price change. Thus, for a given level of investors' uncertainty and diversity of opinion about the cash flow implications of the change, the perceived benefits from costly information processing should increase with the absolute value of the change.⁶ This implies that the number of investors deciding to process underlying market rate/price change information also increases with the absolute value of the rate/price change. Therefore, applying the KV model to signals about changes in interest rates, foreign currency exchange rates, and commodity prices, our first hypothesis, in alternate form, is:

H1: In the absence of FRR No. 48 information, trading volume is positively associated with the absolute value of underlying market rate/price changes.

We hypothesize that after FRR No. 48 information is available, investors' assessments of firms' exposures to underlying market rate/price changes are more precise and homogeneous than was the case in the pre-FRR No. 48 period. In the KV model, both the increase in precision and the reduction in diversity of opinion are likely to reduce the perceived benefits from information processing for any given level of underlying market rate/price change. Thus, the number of investors choosing to engage in costly information processing and trade is likely to be lower and our second hypothesis, in alternate form, is:

H2: Trading volume sensitivity to absolute changes in underlying market rates/prices is lower after FRR No. 48 disclosures than before the disclosures.

Rationale for Focusing on Trading Volume

The KV model implies that investors' uncertainty and diversity of opinion can affect price changes and measures of market liquidity, such as bid-ask spreads, in addition to trading volume. Moreover, Barron et al. (1998) suggest using properties of analysts' forecasts to examine the effects of disclosure on investors' uncertainty and diversity of opinion.

Our empirical tests focus on trading volume rather than these other metrics for the following reasons. First, prior literature (Bamber and Cheon 1995; Barron 1995; Kandel and Pearson 1995; Bamber et al. 1997, 1999) documents reliable associations between trading volume and investors' uncertainty and diversity of opinion (also called disagreement). This literature also documents that disagreement and/or the trading volume related to disagreement can occur in the absence of price changes. For example, Bamber and Cheon

⁶ We examine the absolute value of underlying market rate/price changes because we expect both positive and negative movements of a given magnitude to generate similar perceived information-processing opportunities.

(1995) find that earnings announcements stimulating higher trading volume relative to the absolute magnitude of price change are associated with greater divergence in analysts' pre-disclosure earnings forecasts. These findings suggest that investors' uncertainty and/or diversity of opinion are likely associated more strongly with volume than with price changes. Second, bid-ask spreads quoted in eighths provide only coarse measures of market liquidity (Callahan et al. 1997).

Third, since analysts do not update earnings forecasts daily, it is infeasible to align earnings forecast revisions with daily changes in underlying market rates/prices, as is required in this research. Finally, underlying market rate/price changes can affect firm value and expected future cash flows without affecting *next* period's earnings. For example, an underlying market rate/price change can alter the likely value of transactions forecasted to occur beyond the current year. Similarly, changes affecting the fair value of financial instruments accounted for at historical cost should not prompt near-term earnings forecast revisions unless analysts expect the firm to sell the instruments in the near term. These properties of the current accounting model further diminish the usefulness of analysts' forecasts in our setting.

For these reasons, we believe that trading volume sensitivity to underlying market rate/price changes is the most appropriate metric for testing our hypotheses. We nonetheless recognize that trading volume is a noisy proxy for uncertainty and diversity of opinion. Prior research reveals that proxies for disagreement are significantly associated with trading volume; however, the explanatory power of these models is low, with R^2 s generally less than 20 percent (e.g. Barron 1995; Bamber et al. 1997, 1999).⁷ Moreover, investors trade for reasons other than investors' uncertainty or diversity of opinion (e.g., portfolio rebalancing, liquidity). To the extent that prior research suggests proxies for these other factors stimulating trade, we control for them empirically, as described in the next section.

V. RESEARCH DESIGN

To test the hypotheses, we examine the relation between absolute changes in underlying market rates or prices and the percentage of firms' shares traded before and after firms first issued the FRR No. 48-mandated disclosures. Equations (1)–(3) below specify the regression models used to test our research hypotheses for three subsamples of firms that provided first-time disclosures of exposures to interest rates, foreign currency exchange rates, and commodity prices, respectively.

$$\begin{aligned} \text{SQVOL}_{it} = & \beta_0 + \beta_1 \text{SQMKTVOL}_t + \beta_2 \text{SQ|RET}|_{it} + \sum_{k=-1}^{k=+1} \beta_{3,k} \text{SQ|}\% \Delta \text{IR}|_{t+k} \\ & + \sum_{k=-1}^{k=+1} \beta_{4,k} \text{POST}_{it} \times \text{SQ|}\% \Delta \text{IR}|_{t+k} + \varepsilon_{it} \end{aligned} \quad (1)$$

$$\begin{aligned} \text{SQVOL}_{it} = & \gamma_0 + \gamma_1 \text{SQMKTVOL}_t + \gamma_2 \text{SQ|RET}|_{it} + \sum_{k=-1}^{k=+1} \gamma_{3,k} \text{SQ|}\% \Delta \text{FX}|_{t+k} \\ & + \sum_{k=-1}^{k=+1} \gamma_{4,k} \text{POST}_{it} \times \text{SQ|}\% \Delta \text{FX}|_{t+k} + \omega_{it} \end{aligned} \quad (2)$$

⁷ The R^2 s in our regression models range from 54.28 percent to 84.33 percent. Our research design, however, differs from prior research because, as described in the next section, we examine daily trading volume sensitivity to underlying market rate/price changes over an extended pre- and post-event period, whereas most prior research examines determinants of trading volume measured over a narrow event period.

$$\begin{aligned}
\text{SQVOL}_{it} = & \delta_0 + \delta_1 \text{SQMKTVOL}_t + \delta_2 \text{SQ|RET|}_{it} + \sum_{k=-1}^{k=+1} \delta_{3,k} \text{SQ|}\% \Delta \text{ENERGY|}_{t+k} \\
& + \sum_{k=-1}^{k=+1} \delta_{4,k} \text{POST}_{it} \times \text{SQ|}\% \Delta \text{ENERGY|}_{t+k} \\
& + \sum_{k=-1}^{k=+1} \delta_{5,k} \text{SQ|}\% \Delta \text{NONENERGY|}_{t+k} \\
& + \sum_{k=-1}^{k=+1} \delta_{6,k} \text{POST}_{it} \times \text{SQ|}\% \Delta \text{NONENERGY|}_{t+k} + v_{it}
\end{aligned} \tag{3}$$

To adjust for skewness, we use square root transformations (SQ) for all continuous variables defined below:⁸

- i, t = firm and time subscripts, respectively;
- SQVOL = square root of (number of shares traded divided by the number of outstanding shares);
- SQMKTVOL = square root of (number of shares traded by Dow Jones Industrial Average firms divided by the number of shares outstanding for such firms);
- SQ|RET| = square root of the absolute value of the stock return;
- SQ|% Δ IR| = square root of the absolute value of the percentage change in long-term (10-year Treasury bond) interest rates;
- SQ|% Δ FX| = square root of absolute value of the percentage change in the value of the U.S. Federal Reserve U.S. dollar-weighted index of foreign currency exchange rates;
- SQ|% Δ ENERGY| = square root of absolute value of the percentage change in the Goldman Sachs Commodity Energy Sector Index times a dummy variable set equal to 1 for firms making a disclosure about an energy-related commodity, 0 otherwise;
- SQ|% Δ NONENERGY| = square root of absolute value of the percentage change in the Goldman Sachs Commodity Non-Energy Sector Index times a dummy variable set equal to 1 for firms making a disclosure about a non-energy-related commodity, 0 otherwise;
- POST = dummy variable equal to 1 on days following firms' initial 10-K filing containing FRR No. 48 disclosures, 0 otherwise; and
- ε, ω, ν = random error terms.

We call the coefficients, β_3 , γ_3 , δ_3 , and δ_5 , "volume sensitivity coefficients" (VSCs) because they measure trading volume sensitivities to absolute percentage changes in underlying market rates/prices. The VSCs proxy for trading volume associated with investors' uncertainty and diversity of opinion about the implications of underlying market rate or

⁸ Trading volume (VOL%) has a positive skewness measure of approximately 12. Logarithmic and square root transformations reduce skewness to approximately 10 and 3.5, respectively, so the square root transformation is preferable. Previous studies (e.g., Bamber et al. 1999) used logarithms to partially correct for positive skewness. We tried this (after adding 1 to avoid taking the log of 0) and found that inferences were unaffected.

price changes for firm value. Hypothesis 1 is that in the absence of FRR No. 48 market risk information, the VSCs are positive.

Coefficients β_4 , γ_4 , δ_4 , and δ_6 measure the shift in the VSCs from the pre-FRR No. 48 to the post-FRR No. 48 period. Hypothesis 2 is that these coefficients are negative, since the disclosures mandated by FRR No. 48 increase the precision of investors' assessments of the cash flow implications of underlying market rate or price changes and decrease the diversity of opinion in information processors' assessments. We test H2 by introducing a dummy variable, POST, which takes the value 1 for firm-days after firms' first 10-K disclosure of FRR No. 48 information, 0 for firm-days before disclosure. We interact POST with the absolute values of changes in underlying market rates/prices to isolate the shift in VSCs in the post-FRR No. 48 period.⁹

To allow for the possibility that trading volume on day $t - 1$ occurs in anticipation of a day t underlying market rate/price change, or that trading volume on day $t + 1$ represents additional reaction to the change that occurred on day t , the estimated regression models include a one-day lag and a one-day lead for each underlying market rate/price change. For brevity, the tables report coefficients aggregated across days $t - 1$, t , and $t + 1$.¹⁰

We use ten-year Treasury bond yields to measure $|\% \Delta IR|$ because we expect the most significant interest rate risk exposures for nonfinancial firms to stem from long-term borrowings and related derivatives. We use the U.S. Federal Reserve's U.S. dollar, trade-weighted index for foreign currency exchange rate changes to measure $|\% \Delta FX|$ because the sample firms are typically large multinationals with broad international operations.

The sample firms provide disclosures relating to a variety of commodity price exposures, such as oil, gas, agricultural products, and precious metals; however, more than half the firms in the commodity sample report exposure to energy-related commodity prices. Therefore, we use two indices in Equation (3): the Goldman Sachs Energy Sector Index ($|\% \Delta ENERGY|$), and the Goldman Sachs Non-Energy Sector Index ($|\% \Delta NONENERGY|$).¹¹ We multiply each index by a dummy variable set equal to 1 if firms indicate exposure to either an energy-related commodity or a nonenergy-related commodity, respectively, 0 otherwise.

Trading volume depends on factors besides investors' reactions to underlying market rate/price changes. We control for two of these factors by including additional regressors in Equations (1)–(3). First, like Bamber et al. (1997, 1999) and Tkac (1999), we control for market-wide trading volume, to abstract from trading due to investors' liquidity and portfolio-rebalancing needs. Second, as Abarbanell et al. (1995) advocate, and following Bamber et al. (1997, 1999), we include the absolute value of the rate of return on the firm's common stock, to control for trading volume stemming from the surprise associated with the *average* investor's belief revision on day t .¹² Although the KV model does not suggest

⁹ We also replaced the square root transformation of the market rate/price change variables with a dummy variable equal to 1 on days when the underlying market rate/price changes exceed the median change, 0 otherwise. Our inferences are unchanged.

¹⁰ An aggregated coefficient is similar to a Dimson-adjusted beta, commonly used to calculate market betas for infrequently traded securities (Dimson 1979).

¹¹ The Goldman Sachs Commodity Indices reflect changes in commodity prices. The energy index constituents are heating oil, unleaded gas, crude oil, and natural gas. The non-energy index constituents include various industrial metals, precious metals, livestock, and agricultural products.

¹² Karpoff (1987) reviews literature indicating that trading volume increases with the absolute value of price changes. However, on average, the strength of this relation is greater when stock prices are increasing than when stock prices are decreasing. To allow for differential volume reactions when stock prices are increasing vs. decreasing, we repeat the analyses after including an additional independent variable that reflects the sign of the contemporaneous stock price change. Inferences are unaffected.

that controlling for either of these variables is necessary, we include them to control for factors that prior empirical research has shown to be correlated with trading volume.¹³ Consistent with this research, we predict a positive relation between $SQVOL$ and both $SQ|RET|$ and $SQMKTVOL$.¹⁴

We estimate Equations (1)–(3) by pooling sample firms. The model estimation period spans 120 trading days before, and 120 trading days after, the release of 10-K reports containing the FRR No. 48 information. (A 90-day pre- and post-period window leads to identical inferences.) The estimation period excludes the 10-K report dates. The $POST$ variable takes the value 1 for the 120 trading days beginning after 10-K filing, 0 otherwise. Firm-days in the estimation period span approximately two calendar years, providing ample variation in underlying market rates and prices to estimate the volume sensitivity coefficients efficiently.¹⁵ We use generalized least squares (GLS) estimation to control for first-order autocorrelation and heteroskedasticity in regression errors.¹⁶

VI. SAMPLE AND DESCRIPTIVE STATISTICS

Data on returns, trading volume, and underlying market rates/prices come from DataStream. DataStream reveals 416 nonfinancial SEC registrants that, because their market capitalization exceeded \$2.5 billion on January 28, 1997, were first subject to FRR No. 48 requirements for fiscal years ending after June 15, 1997. Data on market risk disclosures come from the SEC's EDGAR database. We cannot locate ten of the 416 firms on EDGAR. We eliminate an additional 16 firms that experienced takeovers before 10-K filing dates. Finally, we eliminate 12 firms undergoing structural changes, such as mergers, that can alter market risk exposures during the disclosure period. This sample selection procedure yields 378 nonfinancial firms first subject to FRR No. 48 disclosure requirements in the second half of 1997 and first half of 1998.

For 337 of these 378 firms, we can retrieve share return, share trading volume, and market capitalization data around 10-K filing dates. Of the 337 firms, 222 make quantitative FRR No. 48 disclosures; 115 do not, either because they are not *materially* exposed to any category of market risk we examine (i.e., interest rates, foreign currency exchange rates, commodity prices) or they fail to provide disclosures and do not explain why.

Our primary empirical tests focus on three subsamples of the 222 disclosing firms:

- 1) The *interest rate sample* includes 184 firms that make tabular, sensitivity, or VAR (value-at-risk) disclosures about interest rate risk exposures.
- 2) The *foreign currency exchange rate sample* includes 141 firms that give tabular, sensitivity, or VAR disclosures about foreign currency exchange rate risk exposures.

¹³ To the extent that investors' uncertainty and diversity of opinion prompted by underlying market rate/price changes are associated with simultaneous increases in trading volume and absolute price change, controlling for $SQ|RET|$ biases the tests against finding a significant VSC. We also estimate regression equations excluding the control variables, $SQMKTVOL$ and $SQ|RET|$. Inferences are unaffected.

¹⁴ The sample contains large firms with market capitalization exceeding \$2.5 billion; this limits the variation in firm size. As a result, including the market value of equity as a separate control for firm size, as in previous research (e.g., Bamber et al. 1997), does not affect inferences.

¹⁵ Filing dates fall between July 23, 1997 and June 26, 1998, about 11 calendar months. Thus, the full sample test period spans the 468 trading days corresponding closely to the 1997 and 1998 calendar years. As expected, 10-K filing dates cluster in March 1998, the typical filing month for a calendar year-end firm. Specifically, 17.5 percent of the filing dates fall in 1997, 7.7 percent in January or February 1998, 65.2 percent in March 1998, and the remaining 9.6 percent fall in April–June 1998.

¹⁶ Durbin-Watson and White's Chi-square statistics indicate first-order autocorrelation and heteroskedasticity, respectively. We estimate the regression model coefficients using the PROC MODEL procedure in SAS to perform a full information maximum likelihood (FIML) estimation that uses both a GLS covariance matrix and a first-order autoregressive correction.

- 3) The *commodity price sample* includes 58 firms that disclose commodity price risk exposures.

The sum of the numbers of firms in the three subsamples exceeds 222 because some firms are exposed to multiple market risks. The relative numbers of disclosures in the three exposure categories are similar to those in Roulstone's (1999) sample of 25 companies audited by the same Big 6 firm.

Panel A of Table 1 describes the industry composition of the sample and indicates the types of disclosures across industry groups. The 222 disclosing firms come from a variety of industries other than the intentionally excluded financial services industry. Interest rate and foreign currency exchange rate disclosures occur in all industries; however, firms with less internationally diverse operations (e.g., wholesale, retail, transportation, and communications firms) make fewer foreign currency disclosures. The commodity disclosures cluster in the oil and gas, mining, and utilities sectors where exposures to commodity prices, especially energy prices, are most pronounced. Like Roulstone (1999), we find that most disclosures (297 of 389) are sensitivity analyses or VAR.¹⁷

Panel B of Table 1 provides descriptive statistics on selected variables for the pre- and post-FRR No. 48 periods. All statistics come from the distribution of firm-level averages for each variable, calculated over the 120-day pre- or 120-day post-FRR No. 48 period. Average sample firm trading volume does not differ significantly between the periods.¹⁸ The mean market volume, measured as the average of the daily percentage of shares outstanding traded by the Dow Jones Industrial Average companies, is higher in the post-FRR No. 48 period (0.415) than in the pre-FRR No. 48 period (0.411). The average absolute percentage rate of equity return is similar in the two periods, although the median is higher in the post-period (1.497) than in the pre-period (1.385).

The average percentage change in interest rates is slightly lower in the post-period (0.574) than in the pre-period (0.594). The mean percentage change in foreign currency exchange rates vs. the U.S. dollar is similar in the two periods, although the median percentage change is larger in the pre-period (0.288) than the post-period (0.278). The absolute value of the mean percentage change in the energy sector price index is somewhat higher in the post-period (1.374) than the pre-period (1.229). Finally, the absolute value of the mean percentage change in the non-energy sector price index is higher in the post-period (0.334) than the pre-period (0.303).

Panel C of Table 1 reports correlations among the explanatory variables. For consistency with later analyses, we report the correlations based on the square root of each variable. Although generally statistically significant, the magnitudes of these correlations suggest that pair-wise collinearity among these variables is not a significant problem.¹⁹ To further assess the effects of collinearity, we measure the condition index for each regression.

¹⁷ Elmy et al. (1998) and Roulstone (1999) note inconsistent compliance with the tabular disclosure requirements. We classify a firm as meeting the tabular requirements only if it indicates contract sizes and effective rates/prices by year of maturity. We consider firms that fail to provide at least this level of information as nondisclosers.

¹⁸ We delete 20 firm-day observations for which VOL exceeds 1. For all issues on the NYSE, the 1998 *NYSE Fact Book* reports average daily trading volume of about 0.3 percent. The sample firms are traded more actively, likely because they are larger. The insignificant decline in the level of trading volume on post 10-K filing days suggests a general decrease in trading volume after the 10-K filing for sample firms. However, for the square root form of this variable used in the empirical tests reported in Tables 2 through 4, the magnitude of this decline is much smaller (less than one-third as large). We do not observe a similar decline for the sample of 115 nondisclosing firms.

¹⁹ The stronger correlations among the post-period variables ($\text{POST} \times |\% \Delta \text{IR}|$, $\text{POST} \times |\% \Delta \text{FX}|$, $\text{POST} \times |\% \Delta \text{ENERGY}|$, and $\text{POST} \times |\% \Delta \text{NONENERGY}|$) occur because we set all their values to 0 during the pre-period.

TABLE 1
Descriptive Statistics and Correlation Matrix for Firms Providing FRR No. 48 Market Risk Disclosures

Panel A: Number of Firms Making Disclosure Choices

<i>Industry</i>	<i>Total</i>	<i>Interest Rate Exposure</i>		<i>Foreign Currency Exchange Rate Exposure</i>		<i>Commodity Price Exposure</i>	
		<i>Sensitivity or VAR</i>	<i>Tabular</i>	<i>Sensitivity or VAR</i>	<i>Tabular</i>	<i>Sensitivity or VAR</i>	<i>Tabular</i>
Oil and Gas, Mining, Utilities	44	19	15	9	4	28	1
Consumer Products	26	18	9	16	4	10	1
Chemicals, Pharmaceuticals	29	24	1	24	1	4	0
Industrial Products	65	41	14	42	13	6	2
Transportation, Communication	18	12	6	6	0	2	1
Wholesale and Retail	12	3	6	2	3	0	0
Services	18	12	3	7	3	0	0
Miscellaneous	10	5	2	5	2	2	1
Total	222	134 ^a	56 ^a	111	30	52	6

(Continued on next page)

TABLE 1 (Continued)

Panel B: Descriptive Statistics for Firms Providing FRR No. 48 Market Risk Disclosures ($n = 222$)

Variable	Mean	Median	Standard Deviation	(Pre-Post) ^b	
				t-statistic	Z-statistic
VOL (%) – PRE ^c	0.843	0.403	1.760		
VOL (%) – POST ^c	0.675	0.371	0.994	1.24	0.85
MKTVOL (%) – PRE	0.411	0.407	0.014		
MKTVOL (%) – POST	0.415	0.417	0.013	-3.12	-5.66
RET (%) – PRE	1.516	1.385	0.589		
RET (%) – POST	1.598	1.497	0.586	-1.47	-2.05
%ΔIR (%) – PRE	0.594	0.597	0.029		
%ΔIR (%) – POST	0.574	0.574	0.073	3.76	9.05
%ΔFX (%) – PRE	0.272	0.288	0.041		
%ΔFX (%) – POST	0.276	0.278	0.015	-1.55	7.30
%ΔENERGY (%) – PRE	1.229	1.240	0.106		
%ΔENERGY (%) – POST	1.374	1.384	0.087	-15.75	-14.29
%ΔNONENERGY (%) – PRE	0.303	0.289	0.038		
%ΔNONENERGY (%) – POST	0.334	0.343	0.026	-9.86	-10.88
MVE (\$billions) – PRE	199.99	7.591	32.232		
MVE (\$billions) – POST	226.73	8.173	37.437	-0.81	-0.74

(Continued on next page)

TABLE 1 (Continued)
Panel C: Pearson/Spearman Correlation Matrix (Spearman Correlation in Parentheses)^d

<i>Variable</i>	$ RET $	$ \% \Delta IR $	$ \% \Delta FX $	$ \% \Delta ENERGY $	$ \% \Delta NONENERGY $	$POST \times \% \Delta IR $	$POST \times \% \Delta FX $	$POST \times \% \Delta ENERGY $	$POST \times \% \Delta NONENERGY $
MKTVOL	0.22 (0.15)	0.23 (0.20)	0.08 (0.09)	0.11 (0.05)	0.10 (0.13)	0.09 (0.08)	0.07 (0.06)	0.04 (0.02)	0.03 (0.05)
$ RET $		0.09 (0.07)	0.05 (0.04)	0.04 (0.04)	0.02 (0.04)	0.05 (0.04)	0.04 (0.04)	0.02 (0.03)	0.02 (0.03)
$ \% \Delta IR $			0.24 (0.22)	0.04 (0.05)	0.00 (0.00)	0.38 (0.27)	0.09 (0.06)	-0.00 (-0.01)	-0.01 (-0.02)
$ \% \Delta FX $				-0.03 (-0.06)	0.03 (0.02)	0.11 (0.06)	0.37 (0.25)	0.02 (-0.01)	0.01 (-0.01)
$ \% \Delta ENERGY $					0.03 (0.03)	0.07 (0.07)	0.08 (0.07)	0.45 (0.33)	0.08 (0.08)
$ \% \Delta NONENERGY $						0.03 (0.02)	0.04 (0.02)	0.05 (0.03)	0.48 (0.32)
$POST * \% \Delta IR $							0.78 (0.88)	0.69 (0.84)	0.65 (0.81)
$POST * \% \Delta FX $								0.72 (0.86)	0.68 (0.83)
$POST * \% \Delta ENERGY $									0.67 (0.84)

(Continued on next page)

TABLE 1 (Continued)

^a The sum of the combined sensitivity and VAR and tabular disclosures for the interest rate sample (190) exceeds the total number of firms in the interest rate sample because six firms provide both a sensitivity or VAR disclosure and a tabular disclosure with respect to interest rates.

^b t-statistic (Wilcoxon Z-statistic) relates to the test of difference in means (locations) between the pre- and post-10-K filing dates for the sample of disclosing firms. Figures in bold represent t-statistics and Wilcoxon Z-statistics that are significant at the 5 percent level (two-tailed).

^c PRE and POST denote that the means and medians of the variables apply to trading days before and after firms filed their Forms 10-K containing their initial FRR No. 48 disclosures. The statistics are based on the distribution of firm-level averages for each variable, with the average for each variable calculated over the 120-day pre- or 120-day post-FRR No. 48 filing period.

^d Figures in bold denote that correlation statistics are *not* significant at the 5 percent level (two-tailed). For consistency with the subsequent empirical analysis the correlation matrix reports results based on the square root of each variable.

Variables Definitions:

VOL = number of shares traded divided by the number of outstanding shares;
 MKTVOL = market volume, measured as the number of shares traded for all firms included in the Dow Jones Industrial Index divided by the number of shares outstanding for all such firms;

|RET| = absolute value of the rate of return on equity shares of firm *i*;
 |%ΔIR| = absolute value of the percentage change in 10-year Treasury bond rates;
 |%ΔFX| = absolute value of the percentage change in the value of the U.S. Federal Reserve U.S. dollar-weighted index of foreign currency exchange rates;
 |%ΔENERGY| = absolute value of the percentage change in the Goldman Sachs Commodity Energy Sector Index;
 |%ΔNONENERGY| = absolute value of the percentage change in the Goldman Sachs Commodity Non-Energy Sector Index;
 POST = dummy variable equal to 1 on days following a firm's initial 10-K filing containing FRR No. 48 disclosures, 0 otherwise; and
 MVE = market value of common equity.

The largest condition index is less than 10, well below the critical value of 30 (Belsey et al. 1980).

VII. RESULTS

Table 2 reports regression results for the interest rate, foreign currency exchange rate, and commodity price samples in Panels A, B, and C, respectively. Consistent with the results of prior studies, we find that the coefficients of $SQMKTVOL$ and $SQ|RET|$ are positive and statistically significant in each panel at $p \leq 0.0001$. In interpreting our results, we use a statistical cutoff level of $p \leq 0.05$ to reject the null hypothesis.

Primary Tests

The coefficients, β_3 , γ_3 , δ_3 , and δ_5 in Table 2 are the volume sensitivity coefficients (VSCs), which proxy for trading volume associated with investors' uncertainty and diversity of opinion about the implications of underlying market rate/price changes for firm value. Consistent with H1, Table 2 shows that each of these coefficients is positive and statistically significant at $p \leq 0.0002$. We interpret these findings as evidence that, in the absence of disclosure of FRR No. 48 information, investors' uncertainty and diversity of opinion about the firm-specific valuation implications of underlying market rate/price changes leads to positive trading volume sensitivity to absolute changes in interest rates, foreign currency exchange rates, and energy and non-energy commodity prices.

The coefficients β_4 , γ_4 , δ_4 , and δ_6 measure the shift in the volume sensitivity coefficients from the pre- to the post-FRR No. 48 period. Hypothesis 2 predicts that these coefficients are negative. The results are consistent with H2 for the interest rate and foreign currency exchange rate samples. The post-FRR No. 48 change in the VSC is negative and statistically significant (at $p \leq 0.0001$) in both Panels A and B. Consistent with the KV model, we interpret the decline in trading volume sensitivity to changes in these rates/prices as suggesting that FRR No. 48 disclosures reduce investors' uncertainty and diversity of opinion about the implications of changes in interest rates and foreign currency exchange rates for firm value.

Panel C of Table 2 reports the results of estimating Equation (3) for the 58 firms disclosing commodity price exposures. We separately estimate coefficients for exposures to energy prices (33 firms) and non-energy prices (27 firms).²⁰ As hypothesized, we find a marginal decline in the post-FRR No. 48 VSC for firms exposed to energy price changes ($p = 0.0585$); however, for non-energy firms, the VSC shift is not significantly different from 0 at conventional levels ($p = 0.1448$).

There are at least three plausible reasons for the weaker non-energy commodity sample results. First, the non-energy sample is somewhat smaller (27 firms) than the energy sample (33 firms). Second, the broad, non-energy price index is likely a noisy proxy for the specific commodity price exposures of these firms, especially given that non-energy commodity prices are likely to be weakly correlated (e.g., agricultural products and precious metals). In contrast, the energy price index includes fewer commodities (see note 11), and changes in the various energy prices comprising the index are likely more highly correlated. Third, as noted in Section II, FRR No. 48 does not require (although it encourages) the disclosure of quantitative information regarding underlying commodity price exposures. Most sample firms with commodity disclosures do not voluntarily include these exposures

²⁰ The total number of exposures (60) to energy and non-energy prices exceeds 58 because two firms report exposures to *both* energy- and non-energy-related commodities.

TABLE 2
Generalized Least Squares Regression of Trading Volume on Interest Rate, Foreign Currency Exchange Rate, and Commodity Price Changes Pre- and Post-10-K Filings Containing the First FRR No. 48 Disclosures

Panel A: Interest Rate

$$SQVOL_{it} = \beta_0 + \beta_1 SQMKTVOL_t + \beta_2 SQ|RET|_{it} + \sum_{k=-1}^{k=+1} \beta_{3,k} SQ|\% \Delta IR|_{t+k}$$
$$+ \sum_{k=-1}^{k=+1} \beta_{4,k} POST_{it} \times SQ|\% \Delta IR|_{t+k} + \varepsilon_{it}$$

(184 firms; n = 42,360)

Variable	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	Wald Statistic / (p-value): Test $\beta_3 + \beta_4 = 0$
Intercept	?	0.0121	164.86	0.0001	
SQMKTVOL	+	0.7083	2,518.50	0.0001	
SQ RET	+	0.0811	5,891.10	0.0001	
SQ $\% \Delta IR$	+	0.0340	22.83	0.0001	
POST $\times$ SQ $\% \Delta IR$	—	−0.0332	17.58	0.0001	0.01/(0.92)
Adjusted R ²		78.32%			

Panel B: Foreign Currency Exchange Rate

$$SQVOL_{it} = \gamma_0 + \gamma_1 SQMKTVOL_t + \gamma_2 SQ|RET|_{it} + \sum_{k=-1}^{k=+1} \gamma_{3,k} SQ|\% \Delta FX|_{t+k}$$
$$+ \sum_{k=-1}^{k=+1} \gamma_{4,k} POST_{it} \times SQ|\% \Delta FX|_{t+k} + \omega_{it}$$

(141 firms; n = 32,470)

Variable	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	Wald Statistic / (p-value): Test $\gamma_3 + \gamma_4 = 0$
Intercept	?	0.0096	58.10	0.0001	
SQMKTVOL	+	0.8014	2,347.30	0.0001	
SQ RET	+	0.0907	5,641.80	0.0001	
SQ $\% \Delta FX$	+	0.0795	22.19	0.0001	
POST $\times$ SQ $\% \Delta FX$	—	−0.0913	35.89	0.0001	0.44/(0.51)
Adjusted R ²		78.50%			

(Continued on next page)

TABLE 2 (Continued)

Panel C: Commodity Price

$$\begin{aligned}
 SQVOL_{it} = & \delta_0 + \delta_1 SQMKTVOL_{it} + \delta_2 SQ|RET|_{it} + \sum_{k=-l}^{k=+l} \delta_{3,k} SQ|\% \Delta ENERGY|_{t+k} \\
 & + \sum_{k=-l}^{k=+l} \delta_{4,k} POST_{it} \times SQ|\% \Delta ENERGY|_{t+k} + \sum_{k=-l}^{k=+l} \delta_{5,k} SQ|\% \Delta NONENERGY|_{t+k} \\
 & + \sum_{k=-l}^{k=+l} \delta_{6,k} POST_{it} \times SQ|\% \Delta NONENERGY|_{t+k} + \nu_{it}
 \end{aligned}$$

(58 firms; n = 13,369)

Variable	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	Wald Statistic /(p-value): Test	Wald Statistic /(p-value): Test
					$\delta_3 + \delta_4 = 0$	$\delta_5 + \delta_6 = 0$
Intercept	?	0.0130	84.13	0.0001		
SQMKTVOL	+	0.5910	812.88	0.0001		
SQ RET	+	0.0537	933.76	0.0001		
SQ \% \Delta ENERGY	+	0.0288	13.66	0.0002		
POST \times SQ \% \Delta ENERGY	-	-0.0173	3.58	0.0585	2.14/(0.14)	
SQ \% \Delta NONENERGY	+	0.0845	25.95	0.0001		
POST \times SQ \% \Delta NONENERGY	-	-0.0294	2.13	0.1448		10.13/(0.01)
Adjusted R ²		54.28%				

Wald statistics and p-values are two-tailed.

Variable Definitions:

- i, t = firm and time subscripts, respectively;
 $SQVOL$ = square root of (number of shares traded divided by the number of outstanding shares);
 $SQMKTVOL$ = square root of (number of shares traded in the Dow Jones Industrial average firms divided by the number of shares outstanding for such firms);
 $SQ|RET|$ = square root of the absolute value of the stock return;
 $SQ|\% \Delta IR|$ = square root of the absolute value of the percentage change in long-term (10-year Treasury bond) interest rates;
 $SQ|\% \Delta FX|$ = square root of the absolute value of the percentage change in the value of the U.S. Federal Reserve U.S. dollar-weighted index of foreign currency exchange rates;
 $SQ|\% \Delta ENERGY|$ = square root of (absolute value of the percentage change in the Goldman Sachs Commodity Energy Sector Index) times a dummy variable set equal to 1 for firms making a disclosure with respect to an energy-related commodity, 0 otherwise;
 $SQ|\% \Delta NONENERGY|$ = square root of (absolute value of the percentage change in the Goldman Sachs Commodity Non-Energy Sector Index) times a dummy variable set equal to 1 for firms making a disclosure with respect to a non-energy-related commodity, 0 otherwise; and
 $POST$ = dummy variable equal to 1 on days following firms' initial 10-K filing containing FRR No. 48 disclosures, 0 otherwise.

in their FRR No. 48 disclosures, so 10-K readers obtain incomplete commodity price exposure data. Oil and gas firms in the energy sample are less prone to this omission because SFAS No. 69 (FASB 1982) provides industry-specific rules for reporting underlying oil and gas reserves.

Analysis of Post-FRR No. 48 Volume Sensitivity Coefficients

The results in Table 2 suggest that, in the absence of FRR information, there is a positive trading volume sensitivity to the absolute value of daily changes in interest rates, foreign currency exchange rates, and both energy and non-energy commodity prices. The results also show that trading volume sensitivity of firms exposed to changes in interest rates, foreign currency exchange rates, and energy commodity prices declines after the FRR No. 48 disclosures. An open question remains: How successful is FRR No. 48 in reducing investors' uncertainty and diversity of opinion about the firm-specific implications of underlying market rate/price changes for firm value? To provide evidence on this question, we next compute post-FRR No. 48 disclosure VSCs for each of the four types of market rate/price changes, by summing the VSC coefficient estimated in the absence of FRR No. 48 information with the coefficient estimating the shift in VSC in the post-period. We then test whether each of the four post-FRR No. 48 period coefficients is statistically different from zero.

Panels A, B, and C of Table 2 show that we cannot reject the null hypotheses that $\beta_3 + \beta_4 = 0$, $\gamma_3 + \gamma_4 = 0$, or $\delta_3 + \delta_4 = 0$. Thus, after the FRR No. 48 disclosures, daily trading volume is *no longer* sensitive to changes in interest rates, foreign currency exchange rates, or energy commodity prices, respectively. However, for non-energy price exposures, consistent with the insignificant δ_6 coefficient, the sum of δ_5 and δ_6 is positive and statistically significant ($p = 0.01$), indicating that the VSC is still positive with respect to non-energy prices in the post-FRR No. 48 period. Based on our interpretation of the KV model, these results suggest that the FRR No. 48 disclosures reduce investors' uncertainty and diversity of opinion sufficiently to eliminate the trading volume sensitivity to changes in interest rates, foreign currency exchange rates, and energy prices, but not to non-energy commodity prices, in the post-FRR No. 48 period.

Tests for Differences across Disclosure Types

Section II suggests that investors can glean different information from tabular, sensitivity, and VAR disclosures. For example, tabular disclosures provide sufficient information to determine the full distribution of potential future gains/losses in response to likely market rate/price changes, but the calculations involved in making such determinations are complex. In contrast, sensitivity and VAR disclosures are easier to process because they provide point estimates of potential losses. However, both of these disclosures give only one potential loss based on a single, hypothetical underlying market rate/price change scenario, so they are less informative about potential losses in other scenarios. Thus, the three disclosure formats can affect investors' assessments of firms' market risk exposures differently.²¹

²¹ Hodder and McAnally (2001) contend that one may glean sensitivity estimates from tabular disclosures, but they only consider interest rate exposures of financial items. It is sometimes possible to estimate the duration of financial instruments from tabular disclosures and derive a linear estimate of the sensitivity of their fair value to changes in interest rates. However, such approximations ignore the convexity of bond prices with respect to interest rates and other sources of nonlinearity, such as those introduced by options and prepayment provisions.

To explore this possibility, we augment the regression models to allow for differential post-FRR No. 48 shifts in VSCs for sensitivity and VAR disclosures, vs. tabular disclosures. We do not separate the effects of sensitivity vs. VAR disclosures because there are fewer than 30 VAR disclosures in each exposure category. Also, we do not test for differential effects of alternate disclosure formats for the commodity sample because there are only six tabular commodity disclosures. Thus, we focus on the interest rate and foreign currency exchange rate samples in estimating equations (1') and (2').

$$\begin{aligned} \text{SQVOL}_{it} = & \beta_0 + \beta_1 \text{SQMKTVOL}_t + \beta_2 \text{SQ|RET}|_{it} + \sum_{k=-1}^{k=+1} \beta_{3,k} \text{SQ|}\% \Delta \text{IR}|_{t+k} \\ & + \sum_{k=-1}^{k=+1} \beta_{4,k} [\text{POST}_{it} \times \text{SQ|}\% \Delta \text{IR}|_{t+k} \times \text{SENVAR}_i] \\ & + \sum_{k=-1}^{k=+1} \beta_{5,k} [\text{POST}_{it} \times \text{SQ|}\% \Delta \text{IR}|_{t+k} \times \text{TAB}_i] + \varepsilon_{it} \end{aligned} \quad (1')$$

$$\begin{aligned} \text{SQVOL}_{it} = & \gamma_0 + \gamma_1 \text{SQMKTVOL}_t + \gamma_2 \text{SQ|RET}|_{it} + \sum_{k=-1}^{k=+1} \gamma_{3,k} \text{SQ|}\% \Delta \text{FX}|_{t+k} \\ & + \sum_{k=-1}^{k=+1} \gamma_{4,k} [\text{POST}_{it} \times \text{SQ|}\% \Delta \text{FX}|_{t+k} \times \text{SENVAR}_i] \\ & + \sum_{k=-1}^{k=+1} \gamma_{5,k} [\text{POST}_{it} \times \text{SQ|}\% \Delta \text{FX}|_{t+k} \times \text{TAB}_i] + \omega_{it} \end{aligned} \quad (2')$$

All variables are defined as in Equations (1)–(3) except:

SENVAR = dummy variable equal to 1 for firms providing sensitivity or VAR disclosures, 0 otherwise; and

TAB = dummy variable equal to 1 for firms providing tabular disclosures, 0 otherwise.

The results of this analysis appear in Table 3. Panels A and B report interest rate and foreign currency exchange rate sample results, respectively. As expected, in both panels the coefficients on control variables, SQMKTVOL and SQ|RET|, are positive and statistically significant at $p \leq 0.0001$. Moreover, in both panels, β_3 and γ_3 , the VSCs that capture trading volume sensitivity to underlying market rate/price changes in the absence of FRR No. 48 information, are positive and statistically significant, as expected ($p \leq 0.0001$).

Panel A of Table 3 shows that the VSC declines in the post-FRR No. 48 period for firms disclosing interest rate exposures under the tabular alternative ($p \leq 0.0001$). The related post-FRR No. 48 period VSC, represented by sum $\beta_3 + \beta_5$, is negative but not statistically different from 0 ($p = 0.11$). We find a smaller and statistically marginal downward shift in the VSC in the post-FRR No. 48 period for firms disclosing sensitivity analyses or VAR ($p = 0.0505$). Consistent with this marginal downward shift, the related post-FRR No. 48 period VSC represented by $\beta_3 + \beta_4$ remains positive but again is not statistically significant from 0 ($p = 0.10$).

For the foreign currency exchange rate sample in Panel B of Table 3, we find a downward shift in the VSC in the post-FRR No. 48 period for firms disclosing sensitivity analyses

or VAR ($p \leq 0.0001$). The related post-FRR No. 48 period VSC, represented by the sum $\gamma_3 + \gamma_4$, is negative but not statistically different from 0 at $p = 0.09$. In contrast, we do *not* find a significant downward shift in the VSC in the post-FRR No. 48 period for firms disclosing foreign currency exchange rate exposures under the tabular alternative. The related post-FRR No. 48 period VSC represented by the sum $\gamma_3 + \gamma_5$, is positive and significantly different from 0 ($p = 0.03$).²²

Based on our interpretation of the KV model, these results suggest that tabular disclosures are more effective than sensitivity and VAR disclosures in reducing investors' uncertainty and diversity of opinion with respect to interest rate exposures. In contrast, sensitivity

²² We also estimated Equations (1') and (2') separating the sensitivity and VAR disclosures. Although the tests have low power due to the small number of VAR disclosures, we find evidence that the VAR disclosures are associated with VSC reductions in the post-FRR No. 48 period.

TABLE 3

Generalized Least Squares Regression of Trading Volume on Interest Rate and Foreign Currency Exchange Rate Changes Pre- and Post-10-K Filings Containing the First FRR No. 48 Disclosures, Including Interactive Dummy Variables for Sensitivity/VAR Disclosures and Tabular Disclosures

Panel A: Interest Rate

$$\begin{aligned}
 SQVOL_{it} = & \beta_0 + \beta_1 SQMKTVOL_t + \beta_2 SQ|RET|_{it} + \sum_{k=-1}^{k=+1} \beta_{3,k} SQ|\% \Delta IR|_{t+k} \\
 & + \sum_{k=-1}^{k=+1} \beta_{4,k} [POST_{it} \times SQ|\% \Delta IR|_{t+k} \times SENVAR_i] \\
 & + \sum_{k=-1}^{k=+1} \beta_{5,k} [POST_{it} \times SQ|\% \Delta IR|_{t+k} \times TAB_i] + \varepsilon_{it}
 \end{aligned}$$

(184 firms; $n = 42,360$)

Variable	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	Wald Statistic / (p-value): Test $\beta_3 + \beta_4 = 0$	Wald Statistic / (p-value): Test $\beta_3 + \beta_5 = 0$
Intercept	?	0.0121	165.36	0.0001		
SQMKTVOL	+	0.7081	2,513.90	0.0001		
SQ RET	+	0.0811	5,878.80	0.0001		
SQ \% \Delta IR	+	0.0320	20.34	0.0001		
POST $\times$ SQ \% \Delta IR $\times$ SENVAR	-	-0.0056	3.83	0.0505	2.66/(0.10)	
POST $\times$ SQ \% \Delta IR $\times$ TAB	-	-0.0559	14.90	0.0001		2.56/(0.11)
Adjusted R ²		78.32%				

(Continued on next page)

TABLE 3 (Continued)

Panel B: Foreign Currency Exchange Rate

$$\begin{aligned}
 SQVOL_{it} = & \gamma_0 + \gamma_1 SQMKTVOL_t + \gamma_2 SQ|RET|_{it} + \sum_{k=-1}^{k=+1} \gamma_{3,k} SQ|\% \Delta FX|_{t+k} \\
 & + \sum_{k=-1}^{k=+1} \gamma_{4,k} [POST_{it} \times SQ|\% \Delta FX|_{t+k} \times SENVAR_t] \\
 & + \sum_{k=-1}^{k=+1} \gamma_{5,k} [POST_{it} \times SQ|\% \Delta FX|_{t+k} \times TAB_t] + \omega_{it}
 \end{aligned}$$

(141 firms; $n = 32,470$)

Variable	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	Wald Statistic /(p-value): Test	Wald Statistic /(p-value): Test
					$\gamma_3 - \gamma_4 = 0$	$\gamma_3 + \gamma_5 = 0$
Intercept	?	0.0096	58.32	0.0001		
SQMKTVOL	+	0.8013	2,346.70	0.0001		
SQ RET	+	0.0908	5,513.20	0.0001		
SQ \% \Delta FX	+	0.0791	21.99	0.0001		
POST $\times$ SQ \% \Delta FX $\times$ SENVAR	-	-0.1108	45.49	0.0001	2.95/(0.09)	
POST $\times$ SQ \% \Delta FX $\times$ TAB	-	-0.0157	0.31	0.5796		4.51/(0.03)
Adjusted R ²		78.51%				

Wald statistics and p-values are two-tailed.

Variable Definitions:

SENVAR = dummy variable equal to 1 for firms providing sensitivity or VAR disclosures, 0 otherwise;

TAB = dummy variable equal to 1 for firms providing tabular disclosures, 0 otherwise.

All other variables are as defined in Table 2.

and VAR disclosures are more effective than tabular disclosures in reducing investor uncertainty and diversity of opinion for foreign currency exchange rate exposures. Investors, in general, are likely familiar with using tabular interest rate disclosures because financial institutions have long provided regulators with similar information (sometimes called "gap analysis" because tabular interest rate displays focus on the gap between the duration of financial assets and liabilities). In contrast, regulators have not required similar disclosures to describe foreign currency exchange rate exposures. Thus, simple point estimates are likely more effective than extensive tabular disclosures for reducing investors' uncertainty and diversity of opinion for complex foreign currency exchange rate exposures.

Control-Sample Tests

To rule out alternate explanations for our results, we examine the possibility that the observed downward shifts in the post-FRR No. 48 VSCs are associated *not* with the market risk disclosures, *per se*, but with the disclosure of other information in firms' 10-K reports (Cready and Mynatt 1991; Walther 1997). To determine whether there is any basis for this "10-K effect" interpretation, we repeat the analyses around the 10-K filing dates of two control samples:

- 1) *Disclosure sample/prior year*. This sample consists of the 222 disclosing firms in the year before their first FRR No. 48 disclosures.
- 2) *Nondisclosure sample/current year*. This sample consists of 115 firms that did not provide mandated quantitative market risk information in their 10-K reports, even though they were subject to FRR No. 48's requirements during the first year it was effective.

If information other than the market risk disclosures is driving the results, then we expect to observe downward VSC shifts in the post-10-K filing period for both control samples.

Table 4 reports the results of these analyses. In short, we find no significant declines in the VSCs for either control sample (all $p > 0.23$), and the coefficients of other variables are generally consistent with expectations.²³ We therefore conclude that the downward VSC shifts are attributable to FRR No. 48 information, not to other 10-K information.

Model Specification Check

As a final model specification check, we repeat the regression analysis including POST as an additional, separate, explanatory variable. A significant coefficient on this new intercept-shift variable would reflect a post-FRR No. 48-disclosure shift in trading volume *unrelated to underlying market rate/price changes*, whereas H2 states only that FRR No. 48 should cause a decline in trading volume sensitivity to *absolute changes in underlying market rates/prices*. An important, data-related concern hinders this analysis: the post-period intercept-shift and VSC-shift variables are highly collinear, with Pearson correlation statistics greater than 0.75 for each sample.

The results of this robustness check (untabulated) are as follows. Coefficients of both the post-period intercept-shift and VSC-shift variables are uniformly insignificant for the nondisclosing control sample. In contrast, in the disclosing samples, coefficients of the POST intercept variable are negative and statistically significant for the interest rate and

²³ We still predict that the coefficients of the control variables, SQ[RET] and SQMKTVOL, are positive. However, our predictions for the VSCs (β_3 , γ_3 , δ_3 , and δ_5) vary across control samples and types of exposure. For the disclosure sample/prior year, we predict positive VSCs because we expect that firms' market risk exposures were similar to those in the disclosure year. In the nondisclosure sample, we predict zero VSCs for interest rate or foreign currency exchange rate changes because the absence of disclosure suggests that these firms are not materially exposed to interest rate or foreign currency exchange rate changes. However, we predict either positive or zero VSCs for commodity price changes in the nondisclosure sample, because nondisclosure indicates only that their derivative instruments are not materially exposed to commodity price changes. Recall that because FRR No. 48 encourages but does not require disclosure of underlying commodity positions, it is still possible that nondisclosing firms are exposed to material commodity price changes. If so, and if investors' uncertainty and diversity of opinion about the implications of commodity price changes for the underlying commodity positions exists, then we predict positive VSCs. If not, then we predict zero VSCs. Results in Table 4 are consistent with these predictions, with one exception: The β_3 coefficient for the interest rate disclosure/prior year sample is not statistically different from zero.

TABLE 4
Generalized Least Squares Regression of Trading Volume on Interest Rate, Foreign Currency Exchange Rate, and Commodity Price Changes Pre- and Post-10-K Filings for the Disclosure Sample in the Year Prior to FRR No. 48, and for a Sample of Nondisclosing Firms in the Year FRR No. 48 Became Effective

Panel A: Interest Rate

$$SQVOL_{it} = \beta_0 + \beta_1 SQMKTVOL_t + \beta_2 SQ|RET|_{it} + \sum_{k=-1}^{k=-4} \beta_{3,k} SQ| \%AIR|_{t+k} + \sum_{k=-1}^{k=-4} \beta_{4,k} POST_t \times SQ| \%AIR|_{t+k} + \varepsilon_{it}$$

Variable	Disclosure Sample ^a Prior Year (184 firms; n = 42,484)				Nondisclosure Sample ^b Mandated Disclosure Year (115 firms; n = 26,610)			
	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value
Intercept	?	0.0183	295.17	0.0001	?	0.0169	125.49	0.0005
SQMKTVOL	+	0.6655	2,186.80	0.0001	+	0.7114	1,018.60	0.0001
SQ RET	+	0.0855	5,139.40	0.0001	+	0.0998	8,518.90	0.0001
SQ %AIR	+	-0.0088	1.02	0.3115	0	0.0092	0.52	0.4695
POST × SQ %AIR	?	0.0115	1.43	0.2324	?	0.0059	0.22	0.6400
Adjusted R ²		83.51%				74.71%		

(Continued on next page)

TABLE 4 (Continued)

Panel B: Foreign Currency Exchange Rate

$SQVOL_{L,t} = \gamma_0 + \gamma_1 SQMKTVOL_t + \gamma_2 SQ RET _{L,t} + \sum_{k=-1}^{k=+1} \gamma_{3,k} SQ \% \Delta FX _{L,t+k} + \sum_{k=-1}^{k=+1} \gamma_{4,k} POST_t \times SQ \% \Delta FX _{L,t+k} + \omega_{it}$									
Disclosure Sample ^a Prior Year (141 firms; n = 32,646)					Nondisclosure Sample ^b Mandated Disclosure Year (115 firms; n = 26,610)				
Variable	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	
Intercept	?	0.0156	141.84	0.0001	?	0.0179	120.42	0.0001	
SQMKTVOL	+	0.7541	1,845.80	0.0001	+	0.7242	1,103.60	0.0001	
SQ RET	+	0.0946	3,331.90	0.0001	+	0.0997	8,135.10	0.0001	
SQ % ΔFX	+	0.0839	9.72	0.0018	0	-0.0192	0.75	0.3855	
POST × SQ % ΔFX	?	-0.0198	0.78	0.3771	?	-0.0086	0.20	0.6528	
Adjusted R ²		84.33%				74.71%			

(Continued on next page)

TABLE 4 (Continued)

Panel C: Commodity Price

$$SQVOL_{it} = \delta_0 + \delta_1 SQMKTVOL_{it} + \delta_2 SQ|RET|_{it} + \sum_{k=-1}^{k=+1} \delta_{3,k} SQ|\% \Delta ENERGY|_{it+k} + \sum_{k=-1}^{k=+1} \delta_{4,k} POST_{it} \times SQ|\% \Delta ENERGY|_{it+k} + \sum_{k=-1}^{k=+1} \delta_{5,k} SQ|\% \Delta NONENERGY|_{it+k} + \sum_{k=-1}^{k=+1} \delta_{6,k} POST_{it} \times SQ|\% \Delta NONENERGY|_{it+k} + v_{it}$$

Variable	Disclosure Sample ^a Prior Year (58 firms; n = 13,486)				Nondisclosure Sample ^b Mandated Disclosure Year (115 firms; n = 26,610)			
	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value	Predicted Sign	Coefficient Estimate	Wald Statistic	p-value
Intercept	?	0.0179	150.11	0.0001	?	0.0144	84.07	0.0001
SQMKTVOL	+	0.5057	683.75	0.0001	+	0.6926	938.39	0.0001
SQ RET	+	0.0587	1,053.10	0.0001	+	0.0997	7,596.50	0.0001
SQ \% \Delta ENERGY	+	0.1151	12.30	0.0001	+ / 0	0.0185	3.94	0.0470
POST × SQ \% \Delta ENERGY	?	0.0735	1.08	0.2979	?	-0.0036	0.09	0.7638
SQ \% \Delta NONENERGY	+	0.1114	0.66	0.4157	+ / 0	0.0865	25.24	0.0001
POST × SQ \% \Delta NONENERGY	?	0.1483	0.55	0.4569	?	-0.0178	0.69	0.4053
Adjusted R ²			54.85%				74.74%	

Wald statistics and p-values are two-tailed.

^a The sample consists of companies that made 10-K market risk disclosures in the first year they were mandated. We test for significant changes in volume sensitivity coefficients in the prior year when these companies filed 10-K reports that did not contain FRR No. 48 market risk disclosures.

^b The sample consists of companies that did not make 10-K market risk disclosures in the first year they were mandated. We test for significant changes in volume sensitivity coefficients when they filed 10-K reports in that year.

Variables are as defined in Table 2.

foreign currency samples, but not for either of the commodity samples. In addition, coefficients on interactions between POST and changes in underlying market rates/prices remain negative for the interest rate, foreign currency exchange rate, and energy price samples. However, consistent with the high collinearity between the post-period intercept-shift and VSC-shift variables, we observe a reduction in statistical significance for the VSC-shift variables under this specification. The coefficients on the VSC-shift variables are statistically significant at $p \leq 0.0001$ for the foreign currency sample, but not statistically significant at $p < 0.16$ for the interest rate and commodity samples. Importantly, however, the VSC-shift coefficients remain negative and statistically significant at $p \leq 0.0001$ in both the interest rate and foreign currency subsamples where FRR No. 48 disclosures have the strongest effect—the tabular interest rate disclosers and the sensitivity/VAR foreign currency exchange rate disclosers. Thus, even under this alternative specification, the results continue to suggest that FRR No. 48 disclosures are useful to investors; however, the FRR No. 48 effect manifests itself econometrically in both the post-period intercept-shift and VSC-shift variables for the disclosing firms.

VIII. CONCLUSIONS AND LIMITATIONS

We posit that before FRR No. 48 required firms to disclose information about their exposures to changes in interest rates, foreign currency exchange rates, and commodity prices, investors exhibited uncertainty and diversity of opinion about the likely effects on firm value of changes in these market rates/prices. Drawing on Kim and Verrecchia's (1994) analytical model, we argue that this uncertainty and diversity of opinion stimulated trading in firm's equity securities. Specifically, we hypothesize that trading volume sensitivity to absolute percentage changes in underlying market rates/prices is positive in the absence of FRR No. 48 information (H1). We also hypothesize that this trading volume sensitivity declines after FRR No. 48 information facilitates investors' assessments of firms' market risk exposures (H2).

Consistent with H1, absent FRR No. 48 market risk information, we find positive trading volume sensitivity to absolute changes in interest rates, foreign currency exchange rates, and commodity prices. Consistent with H2, after the disclosure of FRR No. 48 information, we find declines in the trading volume sensitivity to changes in interest rates, foreign currency exchange rates, and energy prices. We also find that tabular disclosures are more effective in reducing trading volume sensitivity to interest rate movements, whereas sensitivity and VAR disclosures are more effective in reducing trading volume sensitivity to foreign currency exchange rate movements. We conjecture that the latter finding is consistent with market participants' prior experience in using tabular disclosures to analyze firms' exposures to interest rate movements. In contrast, sensitivity and VAR point estimates more effectively convey complicated foreign currency exchange rate exposures.

To rule out the possibility that downward shifts in trading volume sensitivity are associated with 10-K disclosures in general, and not FRR No. 48 disclosures, we repeat the analyses for a sample of nondisclosing firms and for the sample of disclosing firms in the year prior to the mandated provision of FRR No. 48 market risk information. We detect no reduction in trading volume sensitivity in either control sample. This suggests that our results are attributable to the FRR No. 48 disclosures and *not* to other information in 10-K filings.

A limitation of the study is that trading volume sensitivity to changes in interest rates, foreign currency exchange rates, or commodity prices may capture phenomena other than trading attributable to investors' uncertainty and diversity of opinion about the firm-specific

value implications of such underlying market rate/price changes. Because the theoretical and empirical determinants of trading volume are not completely understood, our interpretation of the results may be vulnerable to the omission of as-yet-unidentified determinants of trading volume. Finally, the results pertain only to very large firms making their initial FRR No. 48 market risk disclosures in 10-K reports, and may not generalize to other firms or other time periods.

The results suggest two additional avenues for future research. First, future studies could examine disclosures after the first compliance year to determine whether subsequent disclosures, especially those confirming or conflicting with information revealed in the initial disclosures, affect trading volume sensitivity to underlying market rate/price changes. Second, as the financial reporting of derivatives and hedging activities continues to evolve, one could extend the analysis to examine the market effects of SFAS Nos. 133 and 138 disclosures, which report the financial statement impact of actual (as opposed to hypothetical) changes in underlying market rates and prices.

APPENDIX

Within each risk category, FRR No. 48 allows companies to present quantitative market risk information using one (or more) of the following three formats:

1. *Tabular presentation* of fair values and contract terms sufficient to determine market risk-sensitive instruments' future cash flow amounts by expected maturity dates. For example, Kellogg's Company provides the following tabular information with respect to interest rate risk in its 1997 10-K report:

Interest Rate Risk

The Company is exposed to interest rate volatility with regard to future issuances of fixed-rate debt and existing issuances of variable-rate debt. Primary exposures include movements in U.S. Treasury rates, London Interbank Offered Rates (LIBOR), and commercial paper rates. The Company uses interest rate caps, and currency and interest rate swaps, including forward swaps, to reduce interest rate volatility and funding costs associated with certain debt issues, and to achieve a desired proportion of variable- vs. fixed-rate debt, based on current and projected market conditions.

The tables below provide information on the Company's significant debt issues and related hedging instruments at year-end 1997. For foreign-currency-denominated debt, the information is presented in U.S. dollar equivalents. Variable interest rates are based on effective rates or implied forward rates as of year-end 1997. Refer to Note 6 within Notes to Consolidated Financial Statements for further information.

	Significant Debt Issues			
	Principal by Year of Maturity (Millions)			
Debt Characteristics	1998	2001	2004	Fair Value (Millions)
Euro Canadian Dollar fixed rate	\$200.0 6.25%			\$186.6
Euro Dollar fixed rate		\$500.0 6.125%		\$502.6
effective rate ^a		6.400%		

Euro Dollar		\$500.0	\$515.3
fixed rate		6.625%	
effective rate ^a		6.354%	
U.S. commercial paper ^b	\$744.2		\$744.2
Weighted av. variable	5.75%		

^a Effective fixed interest rate paid, as a result of settlement of forward interest rate swap at date of debt issuance.

^b \$400 million of commercial paper classified in long-term debt as of year-end 1997. Refer to Note 6 within Notes to Consolidated Financial Statements for further information.

Interest & Currency Swaps & Caps					
Instrument		Year of Maturity (Millions)			Fair Value (Millions)
Characteristics		1998	2001	2002	
Mixed swap – currency/ interest – pay/receive fixed – hedge of existing debt issue	Notional amt. Pay Receive	\$200.0 US\$/4.629% C\$/6.250%			(\$10.8)
Interest rate swap – pay variable/receive fixed – hedge of existing debt issue	Notional amt. Pay Receive		\$200.0 5.96% 6.40%		1.3
Interest rate swap – pay variable/receive fixed – hedge of existing debt issue ^a	Notional amt. Pay Receive			\$225.0 5.600% 6.354%	1.2
Interest rate forward swap – pay fixed/receive variable – hedge of future debt issue	Notional amt. Pay Receive	\$25.0 5.8125% 5.7730%			(0.1)
Interest rate cap – pay fixed if 30-day C.P. rate rises to strike rate – hedge of U.S. commercial paper ^b	Notional amt. Strike Reference		\$225.0 6.33% 5.65%		(0.4)

^a Under the terms of this swap, if three-month LIBOR falls to 4.71% or below, the swap will expire. At year-end 1997, three-month LIBOR was 5.81%.

^b Under the terms of this cap, if the Federal Reserve AA composite rate on 30-day commercial paper increases to 7.83% or above, the cap will expire. At year-end 1997 the rate was 5.65%.

2. *Sensitivity analyses* describing the possible effect on earnings, cash flows, or fair values from selected hypothetical changes in underlying market rates or prices. For example, Time Warner Inc. disclosed in its 1997 10-K filing that:

[B]ased on Time Warner's variable-rate debt and related interest rate swap contracts outstanding at December 31, 1997, each 25 basis point increase or decrease in the level of interest rates would respectively increase or decrease Time Warner's annual interest expense and related cash payments by approximately \$16 million, including \$6 million related to interest rate swap contracts. Such potential increases or decreases are based on certain simplifying assumptions, including a constant level of variable-rate debt and related interest rate swap contracts during the period and,

for all maturities, an immediate, across-the-board increase or decrease in the level of interest rates with no other subsequent changes for the remainder of the period.

3. *Value-at-risk* (VAR) disclosures expressing the potential loss in earnings, cash flows, or fair values from underlying market rate or price changes with a selected likelihood of occurrence. For example, Ford Motor Company made the following disclosure in its 1997 10-K filing:

The VAR analysis calculates the potential risk, with a 99% confidence level, on firm commitment exposures (cash flows), including the effects of foreign currency derivatives. (Translation exposures were not included in the VAR analysis.) The model assumes currency prices are generally normally distributed and draws volatility data from the currency markets. Estimates of correlations of market factors primarily are drawn from the JP Morgan RiskMetrics™ dataset as of December 31, 1997. Based on the overall Company currency exposure at December 31, 1997, including derivative positions, currency movements are projected to affect pre-tax cash flow by less than \$250 million, with a 99% confidence level.

REFERENCES

- Abarbanell, J., W. Lanen, and R. Verrecchia. 1995. Analysts' forecasts as proxies for investors' beliefs in empirical research. *Journal of Accounting and Economics* 20 (1): 31–60.
- Ahmed, A., A. Beatty, and B. Bettinghaus. 2000. Evidence on the efficacy of interest rate risk disclosures by commercial banks. Working paper, Syracuse University, Syracuse, New York and Pennsylvania State University, University Park, Pennsylvania.
- American Institute of Certified Public Accountants (AICPA). 1994. *Improving Business Reporting—A Customer Focus: Meeting the Information Needs of Investors and Creditors*. New York, NY: AICPA.
- Baliga, W. 1994. SEC examining corporate derivative losses. *Journal of Accountancy* 178 (July): 11.
- Bamber, L., and Y. Cheon. 1995. Differential price and volume reactions to accounting earnings announcements. *The Accounting Review* 70 (July): 417–441.
- , O. Barron, and T. Stober. 1997. Trading volume and different aspects of disagreement coincident with earnings announcements. *The Accounting Review* 72 (October): 575–597.
- , ———, and ———. 1999. Differential interpretations and trading volume. *Journal of Financial and Quantitative Analysis* 34 (3): 369–386.
- Barron, O. 1995. Trading volume and belief revisions that differ among individual analysts. *The Accounting Review* 70 (October): 581–597.
- , O. Kim, S. Lim, and D. Stevens. 1998. Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review* 71 (October): 421–433.
- Barth, M., W. Beaver, and W. Landsman. 1996. Value-relevance of banks' fair value disclosures under SFAS No. 107. *The Accounting Review* 71 (October): 513–537.
- , J. Elliott, and M. Finn. 1999. Market rewards associated with patterns of increasing earnings. *Journal of Accounting Research* 37 (2): 387–413.
- Bartov, E., and G. Bodnar. 1994. Firm valuation, earnings expectations, and the exchange-rate exposure effect. *Journal of Finance* 49 (December): 1755–1785.
- Beckett, P. 1997. SEC economist opposes rule on derivatives: Disclosure won't reduce risk and may mislead investors, official says. *Wall Street Journal* (April 7): A9.
- Belsey, D., E. Kuh, and R. Welsch. 1980. *Regression Diagnostics*. New York, NY: Wiley.
- Callahan, C., C. Lee, and T. Yohn. 1997. Accounting disclosures and information asymmetry: Research opportunities associated with bid-ask spreads. *Accounting Horizons* 13 (December): 50–60.
- Collins, D., and M. Venkatachalam. 1996. Derivatives disclosures and interest sensitivity of commercial banks. Working paper, Stanford University, Stanford, California.

- Cready, W., and P. Mynatt. 1991. The information content of annual reports: A price and trading response analysis. *The Accounting Review* 66 (April): 291–312.
- DeMarzo, P., and D. Duffie. 1991. Corporate financial hedging with proprietary information. *Journal of Economic Theory* 53 (April): 261–286.
- , and ———. 1995. Corporate incentives for hedging and hedge accounting. *Review of Financial Studies* 8 (3): 743–772.
- Dimson, E. 1979. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics* 7 (June): 197–226.
- Elmy, F., L. LeGuyader, and T. Linsmeier. 1998. A review of initial filings under the SEC's new market risk disclosure rules. *Journal of Corporate Accounting & Finance* 9 (4): 33–45.
- Financial Accounting Standards Board (FASB). 1982. *Disclosures about Oil and Gas Producing Activities*. Statement of Financial Accounting Standards No. 69. Norwalk, CT: FASB.
- . 1990. *Disclosures of Information about Financial Instruments with Off-Balance Sheet Risk, and Financial Instruments with Concentrations of Credit Risk*. Statement of Financial Accounting Standards No. 105. Norwalk, CT: FASB.
- . 1991a. *Disclosures about Fair Values of Financial Instruments*. Statement of Financial Accounting Standards No. 107. Norwalk, CT: FASB.
- . 1991b. *Discussion Memorandum: Recognition, and Measurement of Financial Instruments*. Norwalk, CT: FASB.
- . 1991c. *Research Report: Hedge Accounting: An Exploratory Study of the Underlying Issues*. Norwalk, CT: FASB.
- . 1994a. *Accounting by Creditors for Impairment of a Loan*. Statement of Financial Accounting Standards No. 114. Norwalk, CT: FASB.
- . 1994b. *Accounting for Certain Investments in Debt and Equity Securities*. Statement of Financial Accounting Standards No. 115. Norwalk, CT: FASB.
- . 1994c. *Accounting by Creditors for Impairment of a Loan—Income Recognition and Disclosures—An Amendment to SFAS No. 114*. Statement of Financial Accounting Standards No. 118. Norwalk, CT: FASB.
- . 1994d. *Disclosures about Derivative Financial Instruments and Fair Value of Financial Instruments*. Statement of Financial Accounting Standards No. 119. Norwalk, CT: FASB.
- . 1995. *Accounting for Certain Investments Held by Not-for-Profit Organizations*. Statement of Financial Accounting Standards No. 124. Norwalk, CT: FASB.
- . 1996a. *Accounting for Transfers and Servicing of Financial Assets and Extinguishments of Liabilities*. Statement of Financial Accounting Standards No. 125. Norwalk, CT: FASB.
- . 1996b. *Exemption from Certain Required Disclosures about Financial Instruments for Certain Non-Public Entities—An Amendment to SFAS No. 107*. Statement of Financial Accounting Standards No. 126. Norwalk, CT: FASB.
- . 1996c. *Deferral of Effective Date of Certain Provisions of SFAS No. 125*. Statement of Financial Accounting Standards No. 127. Norwalk, CT: FASB.
- . 1997. *Reporting Comprehensive Income*. Statement of Financial Accounting Standards No. 130. Norwalk, CT: FASB.
- . 1998. *Accounting for Derivative Instruments and Hedging Activities*. Statement of Financial Accounting Standards No. 133. Norwalk, CT: FASB.
- . 1999. *Accounting for Derivative Instruments and Hedging Activities: Deferral of the Effective Date of SFAS No. 133*. Statement of Financial Accounting Standards No. 137. Norwalk, CT: FASB.
- . 2000. *Accounting for Certain Derivative Instruments and Certain Hedging Activities: An Amendment of SFAS No. 133*. Statement of Financial Accounting Standards No. 138. Norwalk, Connecticut: FASB.
- Hodder, L., L. Koonce, and M-L. McAnally. 2001. SEC market-risk disclosures: Implications for judgment and decision making. *Accounting Horizons* 15 (March): 49–70.
- , and M-L. McAnally. 2001. SEC market-risk disclosures: Enhancing comparability. *Financial Analysts Journal* 57 (March/April): 62–78.

- Jorion, P. 1990. The exchange-rate exposure of U.S. multinationals. *Journal of Business* 63 (3): 331–345.
- Kandel, E., and N. D. Pearson. 1995. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy* 103 (August): 831–872.
- Karpoff, J. 1986. A theory of trading volume. *Journal of Finance* 41 (December): 1069–1087.
- . 1987. The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis* 22 (March): 109–126.
- Kim, O., and R. Verrecchia. 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17 (January): 41–67.
- Linsmeier, T., and N. Pearson. 1997. Commentary: Quantitative disclosures of market risk in the SEC release. *Accounting Horizons* 11 (March): 107–135.
- Logan, D., and P. Montgomery. 1997. Testimony on behalf of the Treasury Management Association concerning disclosures of accounting policies for derivatives and disclosures of quantitative and qualitative information about market risk inherent in market sensitive instruments before the Subcommittee on Banking, Housing, and Urban Affairs, 105th Congress, 1st session, March 4.
- Loomis, C. 1995. Untangling the derivatives mess. *Fortune* (March 20): 50–68.
- McAnally, M-L. 1996. Banks, risk, and FAS No. 105 disclosures. *Journal of Accounting, Auditing, and Finance* 11 (Summer): 453–496.
- Mian, S. 1996. Evidence on corporate hedging policy. *Journal of Financial and Quantitative Analysis* 31 (3): 419–439.
- Nance, D., C. Smith, and C. Smithson. 1993. On the determinants of corporate hedging. *Journal of Finance* 48 (March): 267–284.
- Rajgopal, S. 1999. Early evidence on the informativeness of the SEC's market risk disclosures: The case of commodity price risk exposure of oil and gas producers. *The Accounting Review* 74 (July): 251–280.
- , and M. Venkatachalam. 2000. Are earnings sensitivity measures risk-relevant? The case of oil price risk for the petroleum refining industry. Working paper, Stanford University, Stanford, California.
- Roulstone, D. 1999. Effect of SEC Financial Reporting Release No. 48 on derivative and market risk disclosures. *Accounting Horizons* 13 (December): 343–363.
- Schrand, C. 1997. The association between stock-price interest rate sensitivity and disclosures about derivative instruments. *The Accounting Review* 72 (January): 87–110.
- Securities and Exchange Commission (SEC). 1995. *Proposed Amendments to Require Disclosure of Accounting Policies for Derivative Financial Instruments and Derivative Commodity Instruments and Disclosure of Qualitative and Quantitative Information about Market Risk Inherent in Derivative Financial Instruments, Other Financial Instruments, and Derivative Commodity Instruments*. Release No. 7250. December 28, 1995. SEC. 61 FR No. 5, January 8, 1996. Washington D.C.: Government Printing Office.
- . 1997. *Disclosure of Accounting Policies for Derivative Financial Instruments and Derivative Commodity Instruments and Disclosure of Quantitative and Qualitative Information about Market Risk Inherent in Derivative Financial Instruments, Other Financial Instruments and Derivative Commodity Instruments*. Final Rule (Amendment to Section 210.4-08(n) of Regulation S-X and to Item 310 of Regulation S-B). Release Nos. 33-7386; 34-38223; IC-22487; FR-48; International Series No. 1047; File No. S7-35-95. January 31, 1997. SEC 62 FR No. 27, February 10, 1997. Washington D.C.: Government Printing Office.
- . 1998. *Staff Report: Review of the First Phase of Filings of Disclosures of Quantitative and Qualitative Information about Market Risk Inherent in Derivative Instruments and Other Financial Instruments*. July. Washington D.C.: Government Printing Office.
- . 2000. *Selective Disclosure and Insider Trading*. Securities Act Release No. 7881, Exchange Act Release No. 43154, Investment Company Act Release No. 24599. August. Washington, D.C.: Government Printing Office.
- Smith, C., and R. Stulz. 1985. The determinants of firms' hedging policies. *Journal of Financial and Quantitative Analysis* 20 (4): 391–405.

- Smithson, C., C. Smith, and D. Wilford. 1995. *Managing Financial Risk: A Guide to Derivative Products, Financial Engineering, and Value Maximization*. New York, NY: Irwin Professional Publishing.
- Thornton, D., and M. Welker. 2000. Impact of mandated market risk disclosures on investor-perceived exposure to commodity prices: The case of oil and gas producers. Working paper, Queen's University, Kingston, Canada.
- Tkac, P. 1999. A trading volume benchmark: Theory and evidence. *Journal of Financial and Quantitative Analysis* 34 (March): 89–114.
- Tufano, P. 1996. Who manages risk? An empirical examination of risk management practices in the gold industry. *Journal of Finance* 51 (September): 1097–1138.
- . 1998. The determinants of stock price exposure: Financial engineering and the gold mining industry. *Journal of Finance* 53 (June): 1015–1052.
- Venkatachalam, M. 1996. Value relevance of banks' derivatives disclosures. *Journal of Accounting and Economics* 22 (Aug-Dec): 327–355.
- Walther, B. 1997. Investor sophistication and market earnings expectations. *Journal of Accounting Research* 35 (Autumn): 157–179.
- Wong, M. 2000. The association between SFAS No. 119 derivatives disclosures and the foreign exchange risk exposure of manufacturing firms. *Journal of Accounting Research* 38 (2): 387–418.

Delegation and Incentive Compensation

Venky Nagar
University of Michigan

ABSTRACT: Top management faces two key organizational design choices: (1) how much authority to delegate to lower-level managers, and (2) how to design incentive compensation to ensure that these managers do not misuse their discretion. Theoretical accounting literature emphasizes that top management makes these two choices jointly, but there is little empirical evidence on this assertion. Capitalizing on a unique database of branch manager practices in retail banks, this study provides some of the first evidence on the joint nature of the delegation and incentive compensation choices for lower-level managers. A simultaneous model of these two choices indicates that high-growth, volatile, and innovative banks delegate more authority to branch managers. In turn, branch managers with more authority receive more incentive-based pay. However, in contrast with principal-agent theory, I find no evidence that the extent of incentive compensation plays a significant role in explaining the extent of delegation.

Keywords: *delegation; job authority; incentive compensation.*

Data Availability: *The Wharton Financial Institutions Center survey is not public due to confidentiality agreements. The bank call report data are publicly available at <http://www.fdic.gov>.*

I. INTRODUCTION

Because top management cannot make all the firm's operational decisions, it faces two critical organizational design choices: (1) how much authority to delegate to lower-level employees, and (2) how to design incentive compensation to ensure that these employees do not misuse their discretion (Milgrom and Roberts 1992, Chapter 16; Stiglitz 1994, Chapter 9; Brickley et al. 1996, Chapter 8). A large body of theoretical accounting research has formalized the joint nature of the firm's delegation and incentive

I thank Stanley Baiman, John Core, Adam Zak Gileski, Pierre Hatay, Clement Har, Chip Hunter, Raffi Indjejikian, Christopher Ittner, William Lanen, David Larcker, Russell Lundholm, DJ Nanda, Canice Prendergast, Madhav Rajan, Sofia Tratz, and especially Ken Gaver (associate editor) and two anonymous referees. Comments from the seminar participants at the University of Michigan, Michigan State University, AAA Management Accounting Conference 1999, and the Big 10 Faculty Consortium are greatly appreciated. I am grateful to the Wharton Financial Institutions Center for providing the data.

*Submitted March 2000
Accepted September 2001*

compensation choices (Melumad and Reichelstein 1987; Melumad et al. 1992; Baiman and Rajan 1995; Bushman et al. 2000). Responsibility accounting, a fundamental concept in management accounting, also stresses the link between delegation and incentive compensation, suggesting that lower-level employees' incentives should be closely related to their responsibilities (Horngren et al. 2000, 195–196).

Prior research, however, provides little empirical evidence on how top management balances the delegation of authority and the extent of incentive compensation awarded to lower-level managers. Holthausen et al. (1995) and Bushman et al. (1995) examine incentive compensation for divisional managers, but they do not control for the authority delegated to these managers. Baiman et al. (1995) incorporate both delegation and compensation choices in their study, but examine these two choices *separately* as two different functions of firm characteristics.

The difficulty in acquiring delegation data has also hampered empirical investigation of the joint nature of firms' delegation and incentive compensation choices. For example, Baiman et al. (1995) can only construct a binary proxy for delegation. I capitalize on a unique cross-sectional database of retail banks with relatively detailed information on delegation and incentive compensation choices for branch managers. This database stems from a survey the Wharton Financial Institutions Center (WFIC) conducted in the mid-1990s to understand how banks structure and manage their retail operations.

I use prior theory to develop a simultaneous model of the retail banks' delegation and incentive compensation choices for branch managers. First, the results indicate that high-growth, volatile, and innovative banks delegate more authority to their branch managers than do more stable banks. Second, I find that the extent of the branch manager's incentive compensation increases with the extent of authority delegated to the manager, after controlling for several firm factors. In fact, firm characteristics such as growth, volatility, and innovation play no incremental role in explaining the extent of incentive compensation, after controlling for the extent of delegation. On the other hand, in contrast with principal-agent theory, I find no evidence that the extent of incentive compensation plays a significant role in explaining the extent of delegation. However, I cannot rule out the possibility that this simply reflects the statistical limitations of analyses based on a relatively small sample of 100 banks.

Prior divisional manager compensation studies' failure to control for the extent of delegation can potentially explain some of their puzzling results. In their study of divisional manager compensation, Baiman et al. (1995) predict a positive association between CEO expertise and divisional manager incentive compensation. They argue that an expert CEO can more effectively interpret the divisional performance measures. This lowers the risk to the divisional manager of bonuses based on divisional performance measures, which in turn reduces the cost to the firm of instituting strong incentive pay. However, in contrast to this prediction, Baiman et al. (1995, 224) find that greater CEO division expertise is associated with less powerful incentive compensation for divisional managers. The delegation perspective provides an explanation for this result: CEOs with divisional expertise are less likely to delegate decisions to divisional managers, thus reducing the need to provide the divisional managers with incentive compensation. Indeed, Baiman et al. (1995, 225) find that expert CEOs are less likely to delegate, but do not incorporate this association in explaining the incentive compensation choice.

The remainder of the paper is organized as follows. Section II provides a theoretical discussion of the firm's delegation and incentive compensation choices, based on prior literature. Section III discusses the data set. Section IV develops the empirical measures and the model specification. Section V discusses the results, and Section VI concludes.

II. THEORY

Information required for operational decisions changes rapidly in firms operating in high-growth and uncertain environments. Prior theory suggests that the costs for top management to acquire this information are high in volatile and high-growth firms; consequently, top management delegates authority to divisional managers who have closer access to the necessary information. By delegating authority to these divisional managers, the firm can use this information effectively, while avoiding the costs of collecting and transmitting all information to top management (Melumad and Reichelstein 1987; Kirby 1987; Jensen and Meckling 1992; Bushman et al. 2000; Prendergast 2000).

Although delegation reduces the cost of information gathering, it may create additional costs if divisional managers misuse their authority. To mitigate this moral hazard problem, firms develop incentive compensation linking divisional manager compensation to divisional performance. However, Prendergast (2000) argues that in rapidly changing environments, divisional performance measures are likely to fluctuate for reasons beyond the divisional manager's control. Principal-agent theory indicates that incentives based on noisy performance measures impose risk on the divisional manager, for which the firm must compensate (Kreps 1990, Chapter 16). This compensation, called the risk premium, is the incentive-related cost of delegation, which top management trades off against the benefits of delegation. Thus, top management makes the delegation choice based on both the incentive compensation choice and the nature of the firm's operating environment:

extent of delegation

$$= f(\text{extent of incentive compensation, firm's operating environment}) \quad (1)$$

Because incentive compensation is a cost of delegation, I expect incentive compensation to have a negative coefficient in Equation (1).

The argument outlined above also suggests that incentive compensation is a function of the level of delegation. Further, holding delegation constant, principal-agent theory posits that the more risky the operating environment, the weaker the incentive compensation (Prendergast 2000; Kreps 1990, Chapter 16). Performance measures that form the basis of incentive compensation are more likely to fluctuate in such an environment for reasons beyond the divisional manager's control, and therefore impose greater risk on the divisional manager. Compensating the divisional manager for this risk is costly to the firm. In equilibrium, the firm lowers this cost by offering a weaker incentive contract. Thus, top management makes incentive compensation decisions based not only on the delegation choice, but also on the nature of the firm's operating environment:

extent of incentive compensation

$$= f(\text{extent of delegation, firms' operating environment}) \quad (2)$$

I expect managers with more authority to receive more incentive compensation, so I expect the delegation coefficient to be positive in Equation (2). Equations (1) and (2) form the basis of this study's simultaneous model of delegation and incentive compensation choices.¹

¹ The relation between incentive compensation and delegation is negative in Equation (1), but positive in Equation (2). Such relations are quite common in simultaneous equation models. For example, in the standard Keynesian macroeconomic model of interest rate and output, interest rate negatively affects output, but output positively affects the interest rate (Barro 1993, 551–553).

Several organizational theories view delegation and incentive compensation as the two key choices facing top management (Milgrom and Roberts 1992, Chapter 16; Stiglitz 1994, Chapter 9; Brickley et al. 1996, Chapter 8). However, organizational design is not limited to these two choices, and thus this study's analysis is a simplified partial equilibrium analysis that takes other choices as given. In particular, like Baiman et al. (1995), I do not examine an important aspect of incentive compensation: the choice of specific performance measures in divisional manager incentive contracts (see Bushman et al. 1995; Brickley et al. 1996, Chapter 8). However, this does not appear to be a serious omission, given this study's exclusive focus on retail banking. Banks' widespread use of branch earnings as a performance measure precludes the need to model explicitly the performance measure choice. For example, Ittner et al. (2001) find that branch earnings are a key performance measure for branch manager bonuses, and Ittner and Larcker (1998) find some evidence that branch earnings capture the branch manager's actions in a timelier manner *relative* to other performance measures such as customer satisfaction. Eighty-nine percent of this study's sample uses branch earnings as a basis for branch manager bonuses (though the data on the proportion of bonus based on branch earnings are not available).

III. DATA

This study uses data from two sources. The first source is the Wharton Financial Institutions Center's (WFIC) 1994 survey of retail banks. This survey provides information on the organizational practices in branch banking, the most prominent retail distribution channel at the time (Bank Administration Institute 1998). The banking industry also experienced a rapidly changing environment in the 1990s, with deregulation and the emergence of other retail financial institutions such as mutual funds. Within this environment, banks pursued various organizational strategies in response to external changes. Therefore, despite a relatively homogeneous production function, the sample provides an interesting cross-section of organizational structures and strategies.

The WFIC survey respondents include senior executives of retail banks and branch managers of the head branches. The survey was completed in two phases: first, the WFIC research team approached the 70 largest U.S. bank holding companies. From this group, the team secured the participation of 47 bank holding companies, yielding a sample of 64 banks (seven bank holding companies provided information on two or more banks). The team then contacted the 265 next-largest bank holding companies by mail. From this group, 64 bank holding companies agreed to participate, yielding 71 more banks, and resulting in a total sample of 135 banks.

The second source of data is the bank call reports filed with the Federal Deposit Insurance Corporation (FDIC). The call reports contain quarterly balance sheets, income statements, and other financial data that regulators use to monitor the banks' financial conditions. To ensure that the banks in the final sample have data available from both sources, I match the banks in the WFIC survey data with those in the call reports by name and city, to the extent possible. The relevant data are all available for only 100 banks, which comprise this study's sample.

The combined assets of the banks in the final sample sum to about 22 percent of the total assets of the approximately 10,000 banks listed in the FDIC database. The sample's 1994 capitalization ratio and average net income as a percentage of assets are 9 percent and 1.26 percent, respectively, indicating that the sample banks are, on average, financially healthy.

Although the WFIC survey contacted several employees per bank, it contacted only one employee per bank for each question in the survey (the Appendix provides the identity of the respondent for each survey question). Thus the survey data provide only one response

per bank, even though banks typically have multiple branches. Though a limitation, research on divisional manager compensation research often uses typical or average division measures (Holthausen et al. 1995, 294).

IV. VARIABLE DEFINITIONS AND MODEL SPECIFICATION

This section first describes the measures in the model of the link between delegation and incentive compensation. It then describes the model. Several measures in the model, such as the extent of delegation of decision rights and the nature of the firm's operating environment, are difficult to measure empirically (Jensen 1983). To address this issue, the empirical constructs are based on prior literature, and when possible, measured with multiple variables. The Appendix provides data-specific details on these measures.

Delegation

To measure the extent of delegation from top management to the branch, I use survey items that explore the amount of branch manager authority relative to top management authority in various branch operations. Specifically, I use the following survey questions (details appear in the Appendix):

- 1) Discretion available to the branch manager in hiring tellers.
- 2) Discretion available to the branch manager in awarding promotions in the branch.
- 3) Discretion available to the branch manager in determining the hours the branch is open.
- 4) Discretion available to the branch manager in changing the process for selling new investment products.

In particular, I choose investment products in the fourth question because it reflects the amount of authority the branch manager has over a complex and important retail product.

To assess whether these items reflect a common underlying delegation construct and to reduce the dimensionality of the measure, I conduct a factor analysis. This factor analysis yields one factor with an eigenvalue greater than unity, on which all four items load with weights greater than 0.45. This factor, described in detail in the Appendix, is:

DELEGATE: The standardized aggregated sum of the branch manager's authority in hiring, promoting, setting hours, and changing selling processes.

Although this is an aggregate measure, prior divisional-manager compensation studies have used coarser measures of delegation. For instance, Baiman et al. (1995, 228) use a binary variable to represent whether a division has control over its core functions.

Incentive Compensation

Baker et al. (1994) argue that incentive compensation is the primary performance-based reward scheme in the firm, and that supervisors identify employees who consistently earn high bonuses as talented and thus promote them to the next level. I use the following WFIC survey question to measure the amount of the branch manager's incentive compensation (details appear in the Appendix):

INCENTIVE: The proportion of the typical branch manager's pay that is bonus-based.

INCENTIVE is a categorical measure of whether the typical magnitude of the bonus lies within certain ranges, e.g., "1–6 percent of total pay," or "7–10 percent," etc., with the highest category being "25 percent and above." Theory suggests that managers with

more authority have the potential to earn more incentive compensation. Since the WFIC survey question does not specify whether respondents should indicate potential or actual compensation, INCENTIVE may be a noisy indicator of performance compensation. However, previous research suggests that expected and actual bonuses are very close (Holthausen et al. 1995; Baiman et al. 1995, note 17).

Firm's Operating Environment

To measure the extent to which the firm operates in rapidly changing and uncertain environments, I use three measures: firm growth, volatility of earnings, and the firm's strategic emphasis on innovation. Theory suggests that these factors affect the extent to which top management delegates. If the firm is growing or changing rapidly, then centralized decision making requires top management to track information continuously, which is costly. Delegation of authority to branch managers, who typically have such information due to close, ongoing contacts with customers, becomes an attractive alternative (Baiman and Rajan 1995; Kaplan and Atkinson 1998, Chapter 7). Demsetz (1992) and Brickley et al. (1996, Chapter 8) argue that innovative firms emphasizing new products and new market segments require information such as customer preferences, local opportunities, and lower-level employee adaptability to new products. The intangible nature of this information increases its transmission and interpretation costs across the firm, making delegation an attractive alternative. This study therefore includes the three firm characteristics, growth, volatility, and strategic emphasis on innovation, in the delegation equation.

In addition to affecting delegation, the three firm characteristics directly affect incentive compensation. Theory suggests that, holding all else equal, greater risk or more noise in performance measures imposes more risk on the divisional manager, and therefore reduces the extent to which the firm will use incentive compensation (Prendergast 2000). Prior studies suggest that high-growth and volatile firms' performance measures fluctuate more for exogenous reasons beyond the manager's control, thus imposing more risk on the divisional manager (e.g., Lambert and Larcker 1987). Furthermore, Holthausen et al. (1995) argue that it is difficult to measure accurately the outcomes of managerial actions in firms that place strategic emphasis on innovation of new products and services. This study therefore includes the three firm characteristics in the incentive compensation equation as well.

I next turn to operational measures of growth, volatility, and innovation. To measure firm growth, I use insured deposit growth, since the American Bankers Association (1998) survey indicates that two out of three customers initiate banking relationships with deposits. Given that the WFIC survey took place in December 1994, I measure growth from June 1994 through June 1995 to cover growth prior to December 1994 and anticipated future growth. The firm growth measure is:

GROWTH: The growth in insured deposits from June 1994–June 1995.

Because earnings comprise one of the most important performance measures for branch manager compensation (Ittner et al. 2001), I use the volatility of earnings as a measure of volatility facing the branch manager. Due to unavailability of individual branch earnings data, I use overall bank earnings (scaled by assets to control for heteroskedasticity) to compute volatility. This measure is:

STDROA: Standard deviation of the bank's net income scaled by assets from 1990–1994.²

² Holthausen et al. (1995) also use a five-year period to compute the standard deviation of ROA.

To measure innovation, I use the WFIC survey item on the firm's strategic emphasis on innovation. This survey measure is:

INNOV: The extent to which the firm emphasizes innovation in the design and delivery of products and services.

Finally, since firm size is associated with all critical organizational design choices, I control for the bank's retail division size, using the log of deposits:

SIZE: The log of insured deposits.³

Exogenous Variable for the Delegation Equation

To ensure identification, the simultaneous estimation of Equations (1) and (2) requires exogenous variables for each equation. Although such variables should directly affect only the relevant dependent variable, one can often make a plausible case that a given exogenous variable directly affects all dependent variables (Ittner and Larcker 2001). To mitigate this problem, this study's selection of identifying variables is grounded in prior literature.

I use the number of acquisitions the bank made as the exogenous variable affecting delegation. Acquisitions should reduce delegation, as a successful acquisition requires top-level streamlining and integration of systems, employees, and customers (Ashkenas et al. 1998; Darnell 1999; Carey 2000). Such control can curtail lower-level employees' decision rights.⁴ In contrast to this direct effect on delegation, acquisitions are unlikely to affect the risk or noise in performance measures, especially after controlling for firm size, growth, volatility, and strategy. Thus, acquisitions should affect incentive compensation only indirectly, through the delegation effect.⁵

There is typically a delay between the announcement of an acquisition and its occurrence. To allow for both past and anticipated acquisitions, I count the number of acquisitions the bank made from June 1994 through June 1995. I also control for the size of the acquisitions:

ACQUIRE: The sum of the number of acquisitions the bank made from June 1994 through June 1995, adjusted for size. Following the FDIC categorization, if a particular acquisition increases assets by less than 25 percent, then it is counted as 1; otherwise, it is counted as 2. I then sum all the bank's acquisitions over the June 1994–June 1995 period.

Thus, I estimate the following empirical model for the delegation choice (predicted signs in parentheses):

³ The log of deposits is correlated with log of total assets (commercial plus retail) at 0.98.

⁴ Prior research finds that the combined bank reduces small-business lending after an acquisition, and that one cannot attribute these reductions to the acquirer cutting out bad or negative net present value loans of the acquired bank (Berger et al. 1998; Sapienza 2002). Since small-business lending is largely the responsibility of lower-level managers, the post-acquisition lending contraction suggests a reduction in these managers' job authority (Stein 2002).

⁵ For example, when Travelers, which owned Salomon Smith Barney, merged with Citibank, Salomon Smith Barney bankers' concerns about their bonuses arose primarily from fears of their loss of autonomy in making investment banking deals (Kahn 1998).

$$\begin{aligned}
 \text{DELEGATE} = & \overset{(-)}{a_1} + \overset{(+)}{a_2} \text{ INCENTIVE} + \overset{(+)}{a_3} \text{ GROWTH} + a_4 \text{ STDROA} \\
 & + \overset{(+)}{a_5} \text{ INNOV} + \overset{(?)}{a_6} \text{ SIZE} + \overset{(-)}{a_7} \text{ ACQUIRE} + \varepsilon_1.
 \end{aligned} \tag{3}$$

In this equation, this first regressor, INCENTIVE, should have a negative coefficient as stronger incentives imply a higher risk premium. This premium makes delegation more costly for the firm and, *ceteris paribus*, lowers the equilibrium level of delegation (Prendergast 2000). I expect the next three coefficients on the right side of the equation to be positive, as greater growth, volatility, and innovation require more delegation. I do not predict the direction of the size effect, as firm size can proxy for several constructs.⁶ Finally, I expect the acquisition coefficient to be negative, as acquisitions require more centralized decision making.

Exogenous Variables for the Incentive Compensation Equation

I use proxies for the branch manager's risk aversion as exogenous variables for the incentive Equation (2). Since incentive-related uncertainties are more costly for divisional managers with higher risk aversion, the principal-agent model predicts that, *ceteris paribus*, firms offer weaker incentive contracts to divisional managers who are more risk-averse. Furthermore, Kreps (1990, Chapter 16) suggests that the divisional manager's risk aversion does not directly affect the delegation choice in Equation (1), because the delegation costs to the firm arise not from the divisional manager's risk characteristics, *per se*, but from the payment of incentive compensation wages, which I include as a regressor in the delegation equation.

Prior research finds that the manager's education and experience are associated with the manager's risk aversion. Examining the mutual fund industry, Chevalier and Ellison (1999a, 1999b) find that the risks managers take in their operating decisions increase with manager education and age (a proxy for experience). I, therefore, measure the branch manager's risk aversion by his education and experience with other banks. Educated and experienced branch managers are likely to be wealthier and more confident of their abilities, and thus more tolerant of the financial risk in performance-based rewards. The specific risk aversion measures, described in detail in the Appendix, are:

EDUCATE: Level of the branch manager's education.

EXPERIENCE: A binary variable that measures whether the bank manager has worked for other banks.⁷

If educated branch managers are awarded more decision rights, then one could argue that EDUCATE also belongs in the delegation Equation (1). However, since such managers

⁶ One can argue that large firms are more likely to delegate, but this argument is more valid for large conglomerates with diverse divisions rather than large focused firms. Thus, in this argument, size is a proxy for corporate diversification.

⁷ The WFIC survey measures EDUCATE and EXPERIENCE only for the head branch, which can be problematic if the head branch is not representative of other branches. I find evidence from the WFIC survey on two features of the head branch that suggest that the head branch is similar to other branches. First, the head branch is not different in size (measured by the number of tellers) from the average branch at the 10 percent two-tailed level. Second, the head branch manager's incentive compensation is not significantly different from that of the average branch manager at the 10 percent two-tailed level. Furthermore, this study focuses on *retail* branches, and it is unlikely that retail or small-business customers are very different at the head branch (large commercial customers are not handled at the retail level).

demand adequate compensation for their talent, and incentive compensation is a regressor in the delegation equation, the effect of education on delegation is to some extent already reflected in the delegation equation. Section V documents the robustness of the results to this specification assumption by including EDUCATE as an additional regressor in the delegation equation.

In sum, including the exogenous variables leads to the following incentive compensation equation (predicted signs in parentheses):

$$\begin{aligned} \text{INCENTIVE} = & b_1 + b_2 \overset{(+)}{\text{DELEGATE}} + b_3 \overset{(-)}{\text{GROWTH}} + b_4 \overset{(-)}{\text{STDROA}} + b_5 \overset{(-)}{\text{INNOV}} \\ & + b_6 \overset{(?)}{\text{SIZE}} + b_7 \overset{(+)}{\text{EDUCATE}} + b_8 \overset{(+)}{\text{EXPERIENCE}}_2 \end{aligned} \quad (4)$$

In Equation (4), the first regressor, DELEGATE, should have a positive coefficient, as branch managers who exercise greater authority should receive more incentive compensation. The next three variables should have negative coefficients, as, *ceteris paribus*, high-growth, volatile, and innovative banks are more likely to have noisy performance measures that fluctuate for reasons beyond the branch manager's control. Such noise imposes risk on the branch manager, which the bank responds to by attenuating the extent of incentive compensation. As with the delegation equation, I do not predict the sign of the SIZE coefficient, because firm size can proxy for several constructs. Finally, both EDUCATE and EXPERIENCE should have positive coefficients, as they proxy for risk tolerance, which should be positively related to the extent of incentive compensation.

V. RESULTS

Table 1 presents descriptive statistics for all the measures in the two equations. The survey responses for organizational design measures such as the extent of incentive compensation, the extent of delegation, and the level of firm innovation span the entire set of feasible responses, so these measures have variation. The average incentive compensation comprises 7–10 percent of branch manager's pay.⁸ The mean firm is somewhat innovative, and has a growth rate of 4.83 percent. The mean number of acquisitions is 0.49. A third of the banks in the sample made acquisitions, suggesting that the mean acquisition figure is not driven by a few banks. Finally 44 percent of the branch managers have worked in other banks, and the mean branch manager is college-educated.

Table 2 presents the correlations among the variables. Banks delegating more decision rights to branch managers also use more incentive compensation ($\rho = 0.19$, significant at the 10 percent two-tailed level). As predicted, delegation is significantly higher in high-growth, volatile, and innovative banks ($\rho = 0.18, 0.19$, and 0.27 , respectively, all significant at the 10 percent two-tailed level). However, incentive compensation is not significantly associated with firm growth, volatility, or innovation.

The results in Table 2 also show that the relations between the endogenous variables and the identifying exogenous variables are as expected. The extent of delegation is significantly negatively related to acquisitions ($\rho = -0.17$), while incentive compensation is significantly positively related to branch manager education and experience ($\rho = 0.36$ and 0.19 , respectively). The identifying exogenous variables are significantly correlated with only their respective endogenous variables and not the other endogenous variable, reinforcing the identifying variable selection.

⁸ For comparison, Baiman et al. (1995, Tables 2 and 3) indicate that incentive compensation for divisional managers in the sample averages about 19 percent of pay.

TABLE 1
Descriptive Statistics for Measures of Branch Manager Delegation, Incentive Compensation, and the Firm's Operating Environment in the Retail Banking Industry

<i>Variable</i>	<i>Mean</i>	<i>Median</i>	<i>Standard Deviation</i>	<i>Minimum</i>	<i>Maximum</i>
DELEGATE components:					
1. Branch manager discretion over tellers hired to work in the branch.	2.61	2	1.57	1	7
2. Branch manager discretion over promotions within the branch.	3.53	4	1.18	1	7
3. Branch manager discretion over hours the branch is open.	4.80	5	1.08	1	7
4. Branch manager discretion over changing the process for selling a new investment product.	5.94	6	1.00	3	7
INCENTIVE	2.97	3	1.34	1	7
GROWTH	4.83%	3.20%	21.29%	-91.5%	81.4%
STDROA	0.32	0.26	0.28	0.02	2.04
INNOV	3.37	4	1.00	1	5
SIZE	14.17	13.76	1.33	10.87	17.86
ACQUIRE	0.49	0	0.80	0	3
EDUCATE	2.89	3	1.09	1	5
EXPERIENCE	0.44	0	0.50	0	1

DELEGATE is the extent of authority delegated to the branch manager. INCENTIVE is the amount of incentive compensation. GROWTH is the annual growth in insured deposits. STDROA is the standard deviation of ROA. INNOV is the firm's strategic emphasis on innovation. SIZE is the log of insured deposits. ACQUIRE is the number of acquisitions the bank makes. EDUCATE is the level of branch manager education. EXPERIENCE is the branch manager's experience with other banks. Higher scores for DELEGATE, INCENTIVE, INNOV, EDUCATE, and EXPERIENCE indicate more delegation, stronger incentives, more strategic emphasis on innovation, higher education, and more experience, respectively. The Appendix provides detailed descriptions of the measures. The number of banks is 100.

Table 3 shows the result of the two-stage least squares (2SLS) estimation of Equations (3) and (4). The first-stage adjusted R^2 is 17 percent for the delegation equation, and 19 percent for the incentive compensation equation, similar to that of Holthausen et al. (1995, 300).⁹

The first results column of Table 3 indicates that high-growth, volatile, and innovative banks are likely to delegate more authority (at the 10 percent two-tailed level), as predicted. The identifying variable ACQUIRE is significantly negatively associated with delegation (coefficient = -0.298, t-statistic = -2.23), consistent with the argument that banks engaging in acquisitions are more likely to centralize operations.

Incentive compensation, however, is not a significant predictor of delegation. Including EDUCATE (and, alternatively, EXPERIENCE) as a regressor in the delegation equation to

⁹ Statistical research shows that low explanatory power in first-stage estimation does not necessarily cause second-stage coefficient estimates to have large sampling variance (Nelson and Startz 1990).

TABLE 2
Pearson Correlations among Measures of Branch Manager Delegation, Incentive Compensation,
and the Firm's Operating Environment in the Retail Banking Industry

	<u>DELEGATE</u>	<u>INCENTIVE</u>	<u>GROWTH</u>	<u>STDROA</u>	<u>INNOV</u>	<u>SIZE</u>	<u>ACQUIRE</u>	<u>EDUCATE</u>
INCENTIVE	0.19*							
GROWTH	0.18*	-0.06						
STDROA	0.19*	0.12	-0.11					
INNOV	0.27***	0.09	0.13	-0.03				
SIZE	0.16	0.25***	0.02	0.21**	0.08			
ACQUIRE	-0.17*	-0.15	0.16	-0.05	0.24***	0.12		
EDUCATE	0.03	0.36***	0.01	0.21***	-0.04	0.29***	0.09	
EXPERIENCE	0.07	0.19*	-0.07	-0.05	-0.02	-0.18*	0.01	0.11

*, **, *** Denotes two-tailed significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

DELEGATE is the extent of authority delegated to the branch manager. INCENTIVE is the amount of incentive compensation. GROWTH is the annual growth in insured deposits. STDROA is the standard deviation of ROA. INNOV is the firm's strategic emphasis on innovation. SIZE is the log of insured deposits. ACQUIRE is the number of acquisitions the bank makes. EDUCATE is the level of branch manager education. EXPERIENCE is the branch manager's experience with other banks. Higher scores for DELEGATE, INCENTIVE, INNOV, EDUCATE, and EXPERIENCE indicate more delegation, stronger incentives, more strategic emphasis on innovation, higher education, and more experience, respectively. The Appendix provides detailed descriptions of the measures. The number of banks is 100.

TABLE 3
Simultaneous Estimation of Delegation and Incentive Compensation for Branch Managers in the Retail Banking Industry

<i>Dependent Variable</i> <i>Regressors</i>	<i>DELEGATE</i>		<i>INCENTIVE</i>	
	<i>Predicted Sign</i>	<i>Coefficient (t-statistic)</i>	<i>Predicted Sign</i>	<i>Coefficient (t-statistic)</i>
INTERCEPT		-2.249** (-2.18)		1.312 (0.74)
DELEGATE ^a			(+)	1.145** (2.44)
INCENTIVE	(-)	0.105 (0.57)		
GROWTH	(+)	0.009* (1.92)	(-)	-0.012 (-1.62)
STDROA	(+)	0.583* (1.75)	(-)	-0.657 (-1.18)
INNOV	(+)	0.287*** (2.92)	(-)	-0.155 (-0.93)
SIZE	(?)	0.064 (0.76)	(?)	0.078 (0.72)
ACQUIRE	(-)	-0.298** (-2.23)		
EDUCATE			(+)	0.401*** (3.33)
EXPERIENCE			(+)	0.240 (0.87)
Adjusted R ² (second stage)		0.16***		0.19***
Adjusted R ² (first stage)		0.17***		0.19***
n		100		100

*, **, *** Denotes two-tailed significance at 10 percent, 5 percent, and 1 percent levels, respectively.

^a Indicates significant endogeneity at the 10 percent level (Hausman test).

DELEGATE is the extent of authority delegated to the branch manager. INCENTIVE is the amount of incentive compensation. GROWTH is the annual growth in insured deposits. STDROA is the standard deviation of ROA. INNOV is the firm's strategic emphasis on innovation. SIZE is the log of insured deposits. ACQUIRE is the number of acquisitions the bank makes. EDUCATE is the level of branch manager education. EXPERIENCE is the branch manager's experience with other banks. Higher scores for DELEGATE, INCENTIVE, INNOV, EDUCATE, and EXPERIENCE indicate more delegation, stronger incentives, more strategic emphasis on innovation, higher education, and more experience, respectively. The Appendix provides detailed descriptions of the measures.

control for any incremental education or experience effect on delegation does not change this result. The insignificance of incentive compensation is inconsistent with principal-agent theory, which predicts that incentive compensation is a major cost of delegation. However, an alternative explanation for the insignificance is that the noise in the first-stage estimation of the incentive compensation predictor, coupled with the relatively small sample size of 100 banks, reduces the power of the test.

Turning to the incentive equation, the extent of delegation is significantly positively associated with the extent of incentive compensation (coefficient = 1.145, *t*-statistic = 2.44). The Hausman endogeneity test shows significant endogeneity at the 10 percent level, thus justifying delegation as an endogenous choice variable. However, firm growth, volatility, and innovation are not significant individually (or jointly—the joint *F*-statistic is also insignificant at the 10 percent level). Since these measures proxy for the effect of risk on incentive compensation, the lack of significance may reflect either low risk variation in the sample or an inability of the measures to capture incentive compensation-related risk.¹⁰ Finally, the identifying variable EDUCATE is highly significant, consistent with the argument that, holding delegation constant, a more educated (and presumably less risk-averse) branch manager will receive more incentive compensation in equilibrium.

In summary, the findings indicate that high-growth, volatile, and innovative banks delegate more authority to branch managers, and managers with more authority receive more incentive-based pay. These results provide some of the first empirical evidence on the link between the delegation and incentive compensation choices for lower-level managers in a firm.

The inferences are robust to several sensitivity tests. First, I repeat the tests using a finer measure of ACQUIRE, coding each merger as 1, 2, 3, or 4, depending on whether the merger increases the assets by 0–25 percent, 25–50 percent, 50–75 percent, or more than 75 percent. The results are virtually identical. Second, I test other measures of bank size. Including the number of branches as a regressor does not change the results. Holthausen et al. (1995) and Baiman et al. (1995) argue that division size also affects incentive compensation. I thus include average branch size (measured as firm size divided by the number of branches) as an additional regressor. This inclusion does not affect the results, nor does the inclusion of alternative growth measures such as loan growth. Finally, prior research argues that earnings and capital ratios affect top management's operating, financing, and accounting choices (Collins et al. 1995; Beatty et al. 1995). However, including ROA and the capital ratio as additional control variables does not change the inferences.

VI. CONCLUSIONS

Two key organizational design choices facing top management are: delegation of authority to lower-level managers, and the provision of incentive compensation to ensure that these employees do not misuse their discretion. This study examines empirically the link between delegation and incentive compensation for branch managers in retail banks. A simultaneous equation model of these two choices indicates that high-growth, volatile, and innovative banks delegate more authority to branch managers. In turn, managers with more authority receive more incentive-based pay. However, inconsistent with principal-agent theory (which argues that incentive compensation is a major cost of delegation), incentive pay does not affect the firm's delegation choice.

¹⁰ Low cross-sectional variation in risk could also explain why the risk-premium effect of incentive compensation is insignificant in the delegation equation.

This study's findings are subject to several important caveats. First, no empirical study can simultaneously model all organizational design choices. For example, this study abstracts from the endogeneity of firm characteristics such as innovation. It also abstracts from other reasons for awarding incentive compensation (e.g., mechanism to hire good managers). Therefore, this study is a partial equilibrium analysis. Second, organizational design variables such as delegation, the nature of the firm's operating environment, and managers' risk aversion are measured with error. Third, this study uses observations from firms in a single industry polled at the same time. Cross-sectional dependencies are therefore a significant concern. Fourth, the isolation of exogenous variables to identify the simultaneous equation is difficult, and the empirical model is likely misspecified.

Despite these limitations, this study makes an important contribution to the empirical incentive compensation literature. Although accounting theory has repeatedly emphasized the interrelated nature of delegation and incentive compensation (Melumad and Reichelstein 1987; Melumad et al. 1992; Milgrom and Roberts 1992, Chapter 16; Baiman and Rajan 1995; Bushman et al. 2000), the extensive empirical incentive compensation literature contains, to the best of my knowledge, no direct evidence on this interrelation. This study provides some of the first empirical evidence on the joint nature of the firm's delegation and incentive compensation choices for lower-level managers.

APPENDIX VARIABLE DEFINITIONS

Endogenous Variables

DELEGATE results from a factor analysis of the following four Wharton Financial Institutions Center's (WFIC) 1994 survey questions to the office of the senior executive of the retail bank:¹¹

1. To what extent do the branch managers have a say in the hiring of tellers in the branch?
(Scale 1 to 7, as described below)
2. To what extent do the branch managers have a say in promotions within the branch?
(Scale 1 to 7, as described below)
3. To what extent do the branch managers have a say in the hours the branch is open?
(Scale 1 to 7, as described below)
4. To what extent do the branch managers have a say in changing the process for selling a new investment product?
(Scale 1 to 7, as described below)

- Scale:
1. Branch decides, senior management has no say.
 2. Branch decides, senior management provides advice but no more.
 3. Branch decides, senior management influences the decision.
 4. Both senior management and branch must approve the decision.
 5. Senior management decides, the branch influences the decision.

¹¹ I conduct the factor analysis with oblique rotation, retaining each factor with an eigenvalue greater than unity. Only one factor with an eigenvalue greater than unity emerges, which I label DELEGATE. All four survey questions load on this factor with weights greater than 0.45, the minimum loading threshold recommended in prior literature (Johnson and Wichern 1992, 433; Ittner and Larcker 1995, 7). To construct the DELEGATE factor score, I standardize each survey item to zero mean and unit variance, and sum these items. Prior research recommends the use of such standardized factor scores over actual factor scores (Grice and Harris 1998; Ittner and Larcker 1995, 7). In any event, the standardized factor score for DELEGATE and the actual factor score from the factor analysis are correlated at 0.98.

6. Senior management decides, the branch provides advice but no more.
7. Senior management decides, the branch has no say.

Note: DELEGATE is reverse-coded so that a higher score implies more delegation.

INCENTIVE is the response to the following WFIC survey question to the office of the senior executive of the retail bank:

For the typical branch manager, what percentage of annual pay is bonus or variable?
(Scale: 1 = 0 percent, 2 = 1–6 percent, 3 = 7–10 percent, 4 = 11–15 percent, 5 = 16–20 percent, 6 = 21–25 percent, 7 = greater than 25 percent.)

Firm's Operating Environment

GROWTH:

Growth in insured deposits from June 1994 through June 1995, computed using the FDIC call reports.¹²

STDROA:

The standard deviation of the annual net income (as a percentage of assets to control for heteroskedasticity) from 1990–1994, computed using the FDIC call reports.

INNOV:

The response to the following WFIC survey question to the office of the senior executive of the retail bank:

Please indicate the extent to which the following statement describes your retail bank:
We are innovative in the way we design and deliver products/services.
(Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neither, 4 = Agree, 5 = Strongly Agree)

SIZE:

Log of the bank's insured deposits of the bank (in \$000), computed using the FDIC call reports.

Exogenous Variable for the Delegation Equation

ACQUIRE:

Number of mergers the bank completed between June 1994 and June 1995, collected from the FDIC call reports. To capture merger magnitude, I follow the FDIC categorization and

¹² FDIC Division of Research and Statistics adjusts growth measures for acquisitions as follows. The computation of the growth measures takes into account the values of the acquired institutions for all prior periods, for those institutions that the bank acquires in a pooling-of-interest merger. In contrast to pooling-of-interest accounting, marking-to-market in purchase accounting leads to discontinuities in asset and liability values of the acquired institution. The resulting growth in the combined institution's assets or liabilities over a period that includes the purchase merger could be due to either real growth or marking-to-market. For purchase-accounting acquisitions, the FDIC data have no adjustments, causing performance measures to be distorted. I cannot use the banks' financial reports to correct the data because these reports are available only at the bank holding company level, and not at the individual bank level—the only source of financial data on the individual banks is the FDIC data. However, the distortion appears to be small in the sample: GROWTH's correlation with ACQUIRE, the number of acquisitions, although positive, is insignificant at the 10 percent two-tailed level in Table 2.

code an acquisition increasing the total assets by < 25 percent as 1, > 25 percent as 2. I then sum the codes for all the bank's acquisitions in this period.

Exogenous Variables for the Incentive Compensation Equation

EDUCATE is the response to the following WFIC survey question to the branch manager of the head branch of the retail bank:

What best describes your level of education?

(Scale: 1 = High School, 2 = Some College, 3 = College Degree, 4 = Some Graduate, 5 = Graduate Degree).

EXPERIENCE is the response to the following WFIC survey question to the branch manager of the head branch of the retail bank:

Have you ever worked for other banks?

(Scale: 1 = Yes, 0 = No).

REFERENCES

- American Bankers Association (ABA). 1998. *Retail Banking: An ABA Survey Report*. Washington, D.C.: ABA.
- Ashkenas, R., L. DeMonaco, and S. Francis. 1998. Making the deal real: How GE Capital integrates acquisitions. *Harvard Business Review* (January-February): 165-176.
- Baiman, S., D. Larcker, and M. Rajan. 1995. Organizational design for business units. *Journal of Accounting Research* 33: 205-229.
- , and M. Rajan. 1995. Centralization, delegation, and shared responsibility in the assignment of capital investment decision rights. *Journal of Accounting Research* 33 (Supplement): 135-164.
- Baker, G., M. Gibbs, and B. Holmstrom. 1994. The wage policy of a firm. *Quarterly Journal of Economics* 109: 921-955.
- Bank Administration Institute (BAI). 1998. *New Paradigms in Retail Banking: The RDS Report on the Future of Retail Banking Delivery—Executive Summary*. Chicago, IL: BAI.
- Barro, R. 1993. *Macroeconomics*. 4th edition. Singapore, PRC: John Wiley & Sons.
- Beatty, A., S. Chamberlain, and J. Magliolo. 1995. Managing financial reports of commercial banks: The influence of taxes, regulatory capital, and earnings. *Journal of Accounting Research* 33: 231-261.
- Berger, A., A. Saunders, J. Scalise, and G. Udell. 1998. The effects of bank mergers and acquisitions on small business lending. *Journal of Financial Economics* 50: 187-229.
- Brickley, J., C. Smith, and J. Zimmerman. 1996. *Organizational Architecture: A Managerial Economics Approach*. Homewood, IL: The Irwin Series in Economics.
- Bushman, R., R. Indjejikian, and A. Smith. 1995. Aggregate performance measures in business unit manager compensation: The role of intrafirm interdependencies. *Journal of Accounting Research* 33 (Supplement): 101-128.
- , ———, and M. Penno. 2000. Private pre-decision information, performance measure congruity and the value of delegation. *Contemporary Accounting Research* 17: 561-587.
- Carey, D. 2000. Lessons from master acquirers: A CEO roundtable on making mergers succeed. *Harvard Business Review* (May-June): 145-154.
- Chevalier, J., and G. Ellison. 1999a. Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance. *The Journal of Finance* 54: 875-899.
- , and ———. 1999b. Career concerns of mutual fund managers. *The Quarterly Journal of Economics* 114: 389-433.
- Collins, J., D. Shackelford, and J. Wahlen. 1995. Bank differences in the coordination of regulatory capital, earnings, and taxes. *Journal of Accounting Research* 33: 263-291.

- Darnell, D. 1999. How NationsBank learned to manage people and change in the merger integration process. *Journal of Retail Banking Services* 21 (Summer): 1–6.
- Demsetz, H. 1992. Comments by Harold Demsetz. In *Main Currents in Contract Economics*, edited by L. Werin, and H. Wijkander, 275–281. Oxford, U.K.: Basil Blackwell Press.
- Grice, J., and R. Harris. 1998. A comparison of regression and loading weights for the computation of factor scores. *Multivariate Behavioral Research* 33: 221–247.
- Holthausen, R., D. Larcker, and R. Sloan. 1995. Business unit innovation and the structure of executive compensation. *Journal of Accounting and Economics* 19: 279–314.
- Horngren, C., G. Foster, and S. Datar. 2000. *Cost Accounting: A Managerial Emphasis*. 10th edition. Upper Saddle River, NJ: Prentice Hall.
- Ittner, C., and D. Larcker. 1995. Total quality management and the choice of information and reward systems. *Journal of Accounting Research* 33 (Supplement): 1–34.
- , and ———. 1998. Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research* 36 (Supplement): 1–35.
- , and ———. 2001. Assessing empirical research in managerial accounting: A value-based management perspective. *Journal of Accounting and Economics* 32: 349–410.
- , ———, and M. Meyer. 2001. Anatomy of a balanced scorecard bonus plan. Working paper, University of Pennsylvania, Philadelphia, PA.
- Jensen, M. C. 1983. Organizational theory and methodology. *The Accounting Review* 58: 319–339.
- , and W. Meckling. 1992. Specific and general knowledge and organizational structure. In *Main Currents in Contract Economics*, edited by L. Werin, and H. Wijkander, 251–274. Oxford, U.K.: Basil Blackwell Press.
- Johnson, R., and D. Wichern. 1992. *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Kahn, J. 1998. Puzzling it out at Citicorp; Commercial and investment bankers try working together. *New York Times* (December 18): C1.
- Kaplan, R., and A. Atkinson. 1998. *Advanced Management Accounting*. 3rd edition. Upper Saddle River, NJ: Prentice Hall.
- Kirby, A. 1987. Discussion of: Centralization versus delegation and the value of communication. *Journal of Accounting Research* 25 (Supplement): 19–21.
- Kreps, D. 1990. *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.
- Lambert, R., and D. Larcker. 1987. An analysis of the use of accounting and market measures of performance in executive compensation contracts. *Journal of Accounting Research* 25: 85–125.
- Melumad, N., and S. Reichelstein. 1987. Centralization versus delegation and the value of communication. *Journal of Accounting Research* 25 (Supplement): 1–18.
- , D. Mookherjee, and S. Reichelstein. 1992. A theory of responsibility centers. *Journal of Accounting and Economics* 15: 445–484.
- Milgrom, P., and J. Roberts. 1992. *Economics, Organization, and Management*. Englewood Cliffs, NJ: Prentice Hall, Inc.
- Nelson, C., and R. Startz. 1990. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of Business* 63: S125–S140.
- Prendergast, C. 2000. What tradeoff of risk and incentives? *American Economic Review* 90 (2): 421–425.
- Sapientza, P. 2002. The effects of banking mergers on loan contracts. *The Journal of Finance* 57: 329–367.
- Stein, J. 2002. Information production and capital allocation: Decentralized vs. hierarchical firms. *The Journal of Finance* (forthcoming).
- Stiglitz, J. 1994. *Whither Socialism?* Cambridge, MA: The MIT Press.

Do Firms Use Restructuring Charge Reversals to Meet Earnings Targets?

Stephen R. Moehrle
University of Missouri–St. Louis

ABSTRACT: Many firms that take restructuring charges reverse a portion of those restructuring charge accruals in a later quarter. These reversals increase net income, often substantially. In this study, I investigate whether restructuring charge reversals are associated with incentives to meet or exceed analysts' forecasts, avoid earnings declines relative to prior-year levels, and avoid losses. I examine both the decision to record a reversal and the amount of the reversal, using a sample of 121 reversals recorded between 1990 and 1999. The results suggest that some firms record reversals to beat analysts' forecasts and to avoid reporting net losses. There is also some evidence that firms record reversals to avoid earnings declines. Overall, the results are consistent with firms using restructuring accrual reversals to manage earnings.

Keywords: *Accrual reversal; earnings management; earnings targets; restructuring charge.*

Data Availability: *The data used in this study are publicly available from the sources listed in the paper.*

I. INTRODUCTION

Dechow and Skinner (2000, 235) argue that academics are “unwilling to believe that earnings management is practiced by most firms or that the earnings management that does exist should necessarily concern investors.” In contrast, practitioners and regulators “often see earnings management as pervasive and problematic—and in need of immediate remedial action” (Dechow and Skinner 2000, 235). For example, the Securities

This research is based in part on my dissertation at Indiana University. I thank the members of my dissertation committee, Walt Blacconiere (Chair), Chris Albright, Daniel Beneish, and Jerry Salamon for their many insightful comments and suggestions. I also thank Mark DeFond, Mary Beth Mohrman, Jennifer Reynolds-Moehrle, K. R. Subramanyam, Pamela Stuerke, Jim Wallace, an anonymous referee, and workshop participants at Baruch College–CUNY, Indiana University, University of Missouri–St. Louis, University of Southern California, and Washington University for their helpful comments. I also gratefully acknowledge the contribution of I/B/E/S International, Inc. for providing earnings per share forecast data, available through the Institutional Brokers Estimate System. These data have been provided as part of a broad academic program to encourage earnings expectations research.

*Submitted June 1999
Accepted August 2001*

and Exchange Commission (SEC) Chairman, Arthur Levitt, has expressed concern that earnings management is causing "an erosion in the quality of earnings, and therefore, the quality of financial reporting" (Levitt 1998).

Levitt (1998) cites restructuring charges as one means managers can use to manage earnings. Restructuring charges represent the current period recognition of the estimated amount of current and future costs associated with restructuring activities (e.g., work-force reductions, the redesign of unprofitable operations, etc.). These charges can be substantial. For example, Dechow et al. (1994) find that restructuring charges average 80 percent of net income before the charge. Walter Schuetz, the Chief Accountant in the SEC's Division of Enforcement, indicated that his staff is observing an increasing propensity for firms to reverse portions of their earlier restructuring charges (Schuetz 1998). In contrast to the initial charges, these reversals increase net income. For example, Lucent Technologies reversed \$409 million from restructuring reserves back to income from 1997 to 1999.

In this study, I investigate whether firms use restructuring charge reversals to manage earnings to meet earnings benchmarks. Reversals of these restructuring accruals are observable, explicit transactions that unambiguously increase net income. Examining restructuring charge reversals permits a direct test of potential earnings management based on an observable transaction, rather than on noisy estimates of unobservable earnings management mechanisms such as discretionary accruals.¹ I investigate a sample of 121 reversals recorded between 1990 and 1999 to test whether managers appear to use restructuring charge reversals to avoid missing analysts' forecasts, to avoid earnings declines, and/or to avoid net losses.

I find that managers are more likely to reverse restructuring charges when pre-reversal earnings fall short of analysts' forecasts or when pre-reversal net income is negative. In addition, I find some evidence that managers are more likely to reverse restructuring accruals when pre-reversal net income declined from the previous year. Furthermore, the amount of the reversal increases in the magnitude of the pre-reversal shortfall from analysts' forecasts and the magnitude of pre-reversal losses.

These results are consistent with analysts', auditors', and regulatory authorities' concerns that some firms use restructuring charge reversals to manage earnings. In particular, the results are consistent with recent evidence suggesting that firms manage earnings to avoid "bad news" earnings surprises and net losses (e.g., Burgstahler and Dichev 1997; Degeorge et al. 1999; Burgstahler and Eames 2001). Libby and Kinney (2000, 384) note that although past research suggests that firms manage earnings to meet targets such as forecasts, there is little evidence regarding firms' use of particular methods to meet the benchmarks. This study provides evidence suggesting that managers use a specific accrual, reversals of restructuring charges, to achieve earnings benchmarks.

The remaining sections of this paper are organized as follows. Section II overviews restructuring charges and describes managers' incentives and opportunities to use restructuring charges to manage earnings. Section III develops the hypotheses; Section IV describes the research design; and Section V presents the results. Section VI summarizes the study.

¹ Dechow et al. (1995) and Holthausen et al. (1995) point out that current techniques for estimating discretionary accruals are crude, leading to significant measurement error in the discretionary accrual estimates. Gaver et al. (1995) and Elgers et al. (2001) demonstrate that when estimates of unmanaged earnings are obtained by "backing-out" discretionary accrual estimates from reported earnings, measurement error in discretionary accrual estimates can lead to spurious correlation that biases empirical tests in favor of finding earnings management.

II. MANAGERS' INCENTIVES TO USE RESTRUCTURING CHARGES AND REVERSALS TO MANAGE EARNINGS

Restructuring-related liabilities and expenses include costs incurred for: (1) employee benefits such as severance and termination benefits; (2) elimination and reduction of product lines; (3) consolidation or relocation of plant facilities; (4) new systems development or acquisition; and (5) retraining employees to use newly developed systems. The charge may also include losses from asset impairments and disposals of assets (Emerging Issues Task Force [EITF] 1994).

The decision to restructure is part of the firm's overall investment strategy. However, managers can exercise discretion over the amount and timing of the restructuring charge (and of any subsequent reversal) despite additional guidance from the Financial Accounting Standards Board (FASB) and its EITF on accounting for restructuring activities.² The types of costs included in restructuring charges are sometimes ambiguous, and restructuring charges usually include estimates of future costs. These uncertainties mean that managers can use their discretion to overstate the current-period charge to create a hidden balance sheet reserve that will increase future earnings.

There are several reasons why managers might use the restructuring charge to manage earnings. First, firms can increase future earnings by recognizing the restructuring charge as part of a "big bath" (Elliott and Hanna 1996), and there is evidence that firms record restructuring charges in years when pre-charge accounting earnings are low (e.g., Brickley and Van Drunen 1990; John et al. 1992). Managers may accelerate future costs into restructuring charges (Gill et al. 1996) because analysts usually treat these charges as nonrecurring (Smith and Lipin 1996). Firms can then use the reserve created by these accelerated costs to increase future earnings. Buffett (1999, 15) argues that the size and timing of restructuring charges are "dictated by the cynical proposition that Wall Street will not mind if earnings fall short by \$5 per share in a given quarter, just as long as this deficiency ensures that quarterly earnings in the future will consistently exceed expectations by five cents per share."

Second, prior research suggests that capital markets react positively, on average, to restructuring activities. Brickley and Van Drunen (1990) and Kross et al. (1996) found that despite the large reduction in accounting earnings and book value from substantial restructuring charges, announcements of restructuring activities and related charges result, on average, in positive two-day abnormal returns of roughly 1 percent. Their evidence is consistent with the view of Robert S. Miller, Chairman of Waste Management, Inc., that "somebody woke up to the fact that if you take something as a restructuring charge, investors will forgive you immediately" (Byrnes and Melcher 1998, 134).

² In January 1995, the EITF issued EITF No. 94-3, *Liability Recognition for Certain Employee Termination Benefits and Other Costs to Exit an Activity (Including Certain Costs Incurred in a Restructuring)*, and the FASB released Statement of Financial Accounting Standards (SFAS) No. 121, *Accounting for the Impairment of Long-Lived Assets and for Long-Lived Assets to be Disposed of*. EITF No. 94-3 requires firms that are taking substantial restructuring charges to disclose detailed restructuring plans. It requires an approved and detailed severance plan, including disclosure of the number and type of employees to be terminated and the severance package to be provided. Regarding the exit of a business activity, EITF No. 94-3 requires that the firm record the charge after disclosing details of the timing and scope of the activity. SFAS No. 121, as it relates to restructuring, requires that firms record an asset impairment loss on discontinued activities if the expected future cash flows from the use and disposition of the assets are less than the current carrying value. The SEC Staff Accounting Bulletin (SAB) No. 100, issued in 1999, codified earlier guidance and provided disclosure examples. Because it is difficult to estimate the amount of severance costs or activity exit costs that the firm will actually incur, managers can still exercise ample discretion despite this authoritative guidance.

Third, prior research suggests that firms do not penalize managers for recording restructuring charges. Dechow et al. (1994) find that CEO compensation does not decrease in response to restructuring charges, on average. Gaver and Gaver (1998) also find that executive compensation is largely shielded from the effects of nonrecurring income-decreasing items. Furthermore, John et al. (1992) find no evidence that restructuring efforts lead to higher rates of involuntary top-management turnover.

This body of evidence suggests that managers are not necessarily penalized for recording restructuring charges. However, prior research suggests that failure to meet earnings benchmarks is associated with disproportionately lower stock prices (Kasznik and McNichols 2001; Kinney et al. 2001; Lopez and Rees 2001; Skinner and Sloan 2001) and managerial compensation (Matsunaga and Park 2001). Thus, managers might overestimate restructuring charges to build reserves they can later reverse to boost income in order to meet earnings benchmarks.³ Even when the reversal is a relatively small percentage of the original charge, the reversal can be important if it increases reported earnings to approach or even exceed an earnings benchmark.⁴

Although restructuring accruals provide managers an opportunity to create a hidden reserve, and anecdotal evidence reported in the financial press suggests that managers take advantage of this opportunity, academic research has provided little evidence that managers use restructurings to manage earnings. Jennings et al. (1998) find that restructuring adjustments actually enhance the usefulness of book values as a basis for valuation, suggesting that any post-restructuring distortions are not large. Bens (2000) finds little evidence of earnings management around restructuring activities, and concludes that the use of restructuring charges to manage earnings may not be as pervasive as the financial press suggests. Atiase et al. (2001) find that firms generate higher earnings and returns after restructurings, but they attribute the improvement primarily to increased operating efficiency rather than to earnings management.

In contrast to prior research on restructuring charges, this study focuses on the subsequent reversal of charges. Evidence on restructuring charge reversals provides a relatively direct test of managers' use of restructuring charges to manage earnings. Reversals are observable, in contrast to noisy estimates of unobservable earnings management mechanisms such as discretionary accruals. Thus, these direct tests are more likely to identify earnings management, especially in cases in which managers need only small adjustments to reach earnings targets. In addition, these tests do not rely on noisy discretionary accrual estimates that can lead to spurious correlation that biases empirical tests in favor of finding earnings management.

III. HYPOTHESES DEVELOPMENT

Incentives to Meet Analysts' Forecasts and to Avoid Earnings Decreases and Losses

Evidence in earlier studies suggests that managers manage earnings: (1) to meet or exceed analysts' forecasts (Degeorge et al. 1999; Libby and Kinney 2000; Brown 2001; Burgstahler and Eames 2001; Richardson et al. 2001); (2) to avoid reporting earnings

³ Firms also reverse other types of accruals, such as those associated with other post-retirement employee benefits (Ip 1996).

⁴ Skinner and Sloan (2001) show that missing an earnings benchmark can lead to precipitous stock price declines. Libby and Kinney (2000) provide experimental evidence that an earnings overstatement identified by auditors is less likely to be corrected if the correction would cause reported earnings to fall below the consensus analyst forecast. Instead, auditors in the experiment predict that managers would successfully argue that the error is immaterial in amount. This finding is evidence of managers' desire to meet forecasts and auditors' propensity to accept the opportunistic behavior.

decreases (Burgstahler 1997; Burgstahler and Dichev 1997; Degeorge et al. 1999); and (3) to avoid reporting losses (Burgstahler 1997; Burgstahler and Dichev 1997; Degeorge et al. 1999).

Brown (2001) finds that over the period 1984 to 1999, the median firm's earnings went from barely missing analysts' estimates per I/B/E/S, to meeting estimates, and then to beating estimates. Burgstahler and Dichev (1997, 101) estimate that "8–12% of firms with small pre-managed earnings declines manipulate earnings to achieve earnings increases, and 30–44% of firms with small pre-managed losses manage earnings to create positive earnings," and conclude that avoiding earnings decreases and net losses is a "pervasive phenomenon."

Recent research suggests that the market appears to reward firms that meet or beat earnings targets. Barth et al. (1999) find a price-earnings multiple premium for firms that report consistently increasing earnings and a disproportionately large stock price decrease when the increasing earnings pattern is broken. This is consistent with DeAngelo et al. (1996), who find negative abnormal stock returns averaging 14 percent in the year that an earnings growth pattern is broken.

Bartov et al. (2001) demonstrate that abnormal returns for a quarter are positively and significantly associated with the earnings surprise relative to analysts' forecasts, even when earnings management and/or expectations management are likely to be present. Lopez and Rees (2001) and Kasznik and McNichols (2001) demonstrate a market valuation premium for firms that meet or beat analysts' forecasts and an extra valuation premium for firms that consistently beat forecasts. Skinner and Sloan (2001), Kasznik and McNichols (2001), and Lopez and Rees (2001) also demonstrate a disproportionate valuation penalty for firms that fail to meet forecasts even by a small amount. Skinner and Sloan (2001) attribute the disproportionate price drop following missed forecasts to investor correction of over-optimism about the earnings prospects of growth stocks.

Matsunaga and Park (2001) find that managers also have compensation-based incentives to meet analysts' forecasts and to avoid earnings declines. Specifically, they find that CEO compensation drops beyond the normal penalty for poor performance when the firm misses analysts' forecasts and/or reports earnings decreases in more than one quarter during the year.

This study tests the prediction that managers are more likely to record reversals of restructuring charges when earnings would otherwise fall short of an earnings target:

H1a: *Ceteris paribus*, reversals will occur more frequently in quarters in which pre-reversal earnings are below analysts' earnings forecasts.

H1b: *Ceteris paribus*, reversals will occur more frequently in quarters in which pre-reversal earnings are below actual earnings reported in the same quarter of the prior year.

H1c: *Ceteris paribus*, reversals will occur more frequently in quarters in which the firm experiences a pre-reversal net loss.

In addition, I expect the magnitude of the reversal to increase proportionally to the amount of the shortfall in pre-reversal earnings:

H2a: *Ceteris paribus*, the greater the pre-reversal earnings shortfall relative to analysts' earnings forecasts, the greater the amount of the reversal.

H2b: *Ceteris paribus*, the greater the pre-reversal earnings shortfall relative to earnings in the same quarter of the prior year, the greater the amount of the reversal.

H2c: *Ceteris paribus*, the greater the pre-reversal net loss, the greater the amount of the reversal.

Political Cost and Contract-Based Incentives

Political costs and debt and compensation contracts based on accounting numbers also provide incentives for earnings management (see Watts and Zimmerman 1986, 1990). The political cost argument suggests that politically visible firms use accounting discretion to decrease reported earnings to avoid wealth outflows arising from allegations of excessive profitability. Because the firms in this study make income-increasing accounting adjustments, political costs cannot explain these reversals.

Contracts based on accounting numbers can provide managers with incentives to manage earnings. First, firms have incentives to avoid violating restrictive debt covenants (i.e., technical default).⁵ As the firm nears debt covenant constraints, managers likely use accounting discretion to create debt covenant slack (Daley and Vigeland 1983; DeFond and Jiambalvo 1994; Sweeney 1994; Dichev and Skinner 2001). Second, compensation contracts can provide managers with incentives to manage earnings. For example, Gaver and Gaver (1998) find some evidence that nonrecurring income-increasing transactions are associated with higher managerial compensation.

Sensitivity analyses reported later suggest that the study's inferences continue to hold after controlling for debt contracting and executive bonus-based incentives for earnings management.

IV. RESEARCH DESIGN

Sample Selection

I used the "Company/Allnews" files in the Lexis/Nexis database to identify firms that recorded reversals and the amounts of the reversals.⁶ This search identified 147 reversals recorded by 114 different firms between 1990 and 1999. Analyses of reversals to avoid missing analysts' forecasts required median consensus analysts' forecasts and actual earnings per share (EPS) amounts from The Institutional Brokers' Estimate System (I/B/E/S). Analyses of reversals to meet prior-year earnings or to avoid net losses required net income and EPS data from Compustat. Of the 4,560 possible firm-quarters over the ten-year period, I eliminated 890 firm-quarters that could not have had reversals because the quarters preceded the restructuring charges. I also excluded firm-quarter observations if earnings were so far below the benchmarks that reversals would likely be insufficient to raise earnings to meet the benchmarks. As a result, I excluded six non-reversal firm-quarter observations from the analyst forecast analysis because EPS was more than \$5 below analyst's forecasts, and I excluded ten non-reversal quarters from the prior-year earnings and loss benchmark analyses because earnings were more than \$5 below prior-year EPS or the firm reported a loss per share of more than \$5. Table 1 summarizes the sample selection, and shows that the final reversal and non-reversal samples for the analysts' forecast error analyses and prior-year earnings/net loss analyses include 83 and 121 reversal quarters, and 1,748 and 2,434 non-reversal quarters, respectively.

⁵ Beneish and Press (1993) discuss the costs of technical default.

⁶ The Company/Allnews file contains press releases issued by publicly traded firms. The search term was "restruct! and revers!".

TABLE 1
Sample Selection for Restructuring Charge Reversal and Non-Reversal Firm-Quarters, 1990–1999

	<i>Analysts' Forecast Benchmark</i>	<i>Prior-Year Earnings and Loss Benchmarks</i>
Total Number of Quarters		
Reversal Firms	114	114
Firm-Quarters (114 firms $\times$ 10 years $\times$ 4 quarters)	<u>4,560</u>	<u>4,560</u>
Reversal Quarters		
Quarters with restructuring charge reversals	147	147
Less: Reversal quarters with insufficient I/B/E/S or Compustat data for determining earnings benchmarks ^a	<u>(64)</u>	<u>(26)</u>
Usable reversal firm-quarter observations	<u>83</u>	<u>121</u>
Non-Reversal Quarters		
Quarters without restructuring charge reversals	4,413	4,413
Less: Quarters preceding restructuring charges	(890)	(890)
Less: Non-reversal quarters with insufficient I/B/E/S or Compustat data for determining earnings benchmarks ^a	(1,769)	(1,079)
Less: Extreme observations ^b	<u>(6)</u>	<u>(10)</u>
Usable non-reversal quarters in final sample	<u>1,748</u>	<u>2,434</u>
Total firm-quarter observations	<u>1,831</u>	<u>2,555</u>

^a I excluded firm-quarter observations from the analysts' forecast analysis if analysts' forecasts or earnings per share data were missing from I/B/E/S. I excluded firm-quarter observations from the prior-year earnings analysis if net income in the current quarter or four quarters prior was missing from Compustat. I excluded firm-quarter observations from the loss avoidance analysis if net income in the current quarter was missing from Compustat.

^b I excluded six firm-quarter observations from the analysts' forecast benchmark analysis because EPS was more than \$5 below analysts' forecasts and I excluded ten firm-quarter observations from the prior-year earnings and loss benchmark analyses because EPS was more than \$5 below prior-year EPS or because EPS was a loss per share of more than \$5.

Empirical Models

I use a Chi-square test of frequency to determine whether firms are more likely to record reversals when pre-reversal earnings fall short of analysts' forecasts (H1a), or same-quarter-prior-year earnings (H1b), or when pre-reversal net losses occur (H1c). I estimate the following ordinary least squares (OLS) regressions to examine the relation between the magnitude of the reversal and the amount of shortfall from analysts' forecasts (H2a), the amount of shortfall from same-quarter-prior-year earnings (H2b), and the amount of pre-reversal net losses (H2c). I adjust all amounts to year 2000 price levels to control for inflation.⁷

⁷ Inflation adjustments are based on the ratio of the Consumer Price Index (CPI) in the year to which the data pertain and the CPI in the year 2000.

$$\text{REVPS}_{i,j} = \alpha_1 + \beta_1 \text{SHORTFALLAF}_{i,j} + \varepsilon \quad (1)$$

where $\text{REVPS}_{i,j}$ (reversal per share) = the amount of the reversal firm i recorded in quarter j , divided by the number of common shares outstanding at the end of quarter j ; and $\text{SHORTFALLAF}_{i,j}$ (earnings shortfall relative to analysts' forecasts) = the last median consensus I/B/E/S analysts' EPS forecast for firm i in quarter j minus pre-reversal actual EPS. I define pre-reversal actual earnings as: post-reversal actual earnings reported by I/B/E/S – reversal per share net of 40 percent tax. If the analyst forecast exceeds pre-reversal EPS, then SHORTFALLAF is positive.⁸

$$\text{REVPS}_{i,j} = \alpha_2 + \beta_2 \text{DECLINE}_{i,j} + \varepsilon' \quad (2)$$

where $\text{DECLINE}_{i,j}$ (earnings decline from prior-year earnings) = actual EPS for firm i in quarter j of the prior year (excluding restructuring charges if any)⁹ minus actual pre-reversal EPS for firm i in quarter j . If the prior-year earnings exceed the current year pre-reversal earnings, then DECLINE is positive.

$$\text{REVPS}_{i,j} = \alpha_3 + \beta_3 (\text{LOSSQTR}_{i,j} \times |\text{LOSSBR}_{i,j}|) + \varepsilon'' \quad (3)$$

where $\text{LOSSQTR}_{i,j}$ (loss quarter) = 1 if firm i has a pre-reversal net loss in quarter j and 0 otherwise; and $\text{LOSSBR}_{i,j}$ (loss per common share before reversal) = the absolute value of the actual pre-reversal loss per share for firm i in quarter j .

If the magnitude of the restructuring charge reversal is increasing in the amount of the earnings shortfall relative to a specific earnings benchmark, then the β associated with the benchmark should be significantly positive.

V. RESULTS

Descriptive Statistics

Table 2 reports inflation-adjusted descriptive statistics for the sample firms in reversal and non-reversal quarters. Non-reversal quarters are slightly larger in terms of median assets, liabilities, and revenues. The pre-reversal median net income is \$1.35 million in reversal quarters, and is considerably lower than the median net income of \$2.27 million in non-reversal quarters. The lower median revenue and net income in reversal quarters is consistent with a need to manage earnings upward. Sample reversals range from \$60 thousand to \$163.23 million with a median of \$2.86 million (\$0.12 per share). Per-share restructuring charge reversals range from \$0.01 to \$2.40.

Panel C of Table 2 shows that firms recorded reversals less frequently in the early 1990s, but that the frequency increased in the middle of the decade and remained high throughout the rest of the sample period. In addition, firms recorded nearly half (48.8

⁸ To conduct the analysts' forecast tests, I used I/B/E/S actual and forecasted earnings to align the definition of the reported earnings measure and the forecast, as Abarbanell and Lehavy (2000) suggest. In some instances, I/B/E/S excludes unusual items from reported earnings (e.g., nonrecurring gains and losses). Abarbanell and Lehavy (2000) discuss this practice. I examined I/B/E/S actual EPS for each reversal firm-quarter to determine if the restructuring charge reversal was excluded. If it was excluded, then I added the reversal back to I/B/E/S EPS before performing the empirical tests.

⁹ When a firm recorded a restructuring charge in the same quarter of the prior year, I compared the current quarter pre-reversal EPS to the pre-restructuring charge EPS in the prior-year quarter. The inferences are unchanged when I omit these observations.

TABLE 2
Descriptive Statistics for Restructuring Charge Reversal Quarters and Non-Reversal Quarters, 1990–1999

Panel A: Reversal Quarters

(in millions of dollars except per share amounts)

	<i>Reversal Quarters</i>			
	<i>Mean</i>	<i>Median</i>	<i>Maximum</i>	<i>Minimum</i>
Total assets	3,049.72	411.40	42,407.35	9.42
Total liabilities	2,144.20	210.50	27,668.47	4.11
Total revenues	790.90	106.24	11,652.48	3.39
Net income before reversal	19.81	1.35	838.42	−652.91
Reversal amount	10.61	2.86	163.23	0.06
Reversal amount per share	0.30	0.12	2.40	0.01

Panel B: Non-Reversal Quarters

(in millions of dollars)

	<i>Non-Reversal Quarters</i>			
	<i>Mean</i>	<i>Median</i>	<i>Maximum</i>	<i>Minimum</i>
Total assets	2,598.86	437.29	49,843.22	0.37
Total liabilities	1,706.39	259.33	36,907.14	3.42
Total revenues	628.29	136.74	12,469.85	0.54
Net income	21.13	2.27	1,466.56	−805.05

Panel C: Sample Reversals by Year 1990–1999

1990: 3
 1991: 5
 1992: 5
 1993: 3
 1994: 19
 1995: 16
 1996: 7
 1997: 16
 1998: 24
 1999: 23

Panel D: Sample Reversals by Quarter

First quarter: 21 (17.4%)
 Second quarter: 17 (14.0%)
 Third quarter: 24 (19.8%)
 Fourth quarter: 59 (48.8%)

percent) of the sample reversals in the fourth quarter. This result is similar to Elliott and Shaw's (1988) result that 63 percent of potentially discretionary write-offs occur in the fourth quarter. As Elliott and Shaw (1988, 97) point out, the clustering of potentially discretionary charges and reversals in the fourth quarter is consistent with either: (1) earnings management or (2) "characteristics of U.S. management practices, such as the auditing of annual results but not quarterly results or an institutionalized planning process which formalizes plans (such as restructurings) on an annual cycle." In this study, I investigate whether the timing and magnitude of reversals are associated with earnings benchmarks, in an effort to ascertain whether these potentially discretionary reversals reflect earnings management.

Table 3 reports descriptive statistics associated with earnings and earnings benchmarks. Median pre-reversal EPS (per Compustat) were \$0.06 and \$0.12 for reversal and non-reversal quarters, respectively. This evidence is consistent with lower pre-reversal earnings in reversal quarters ($p < 0.02$, one-tailed test). There is no significant difference in analysts' forecasts in reversal and non-reversal quarters. This suggests that earnings are lower in reversal quarters than in non-reversal quarters relative to analysts' expectations. Specifically, the median pre-reversal EPS was \$0.02 less than analysts' forecasts in reversal quarters, but median EPS was equal to analysts' forecasts in non-reversal quarters. This difference is significant at $p < 0.01$ (one-tailed), and suggests that firms use reversals to increase earnings that would otherwise fall below analysts' forecasts.¹⁰ Median pre-reversal EPS is \$0.02 less than median same-quarter-of-the-prior-year EPS in reversal quarters, and equal

¹⁰ In Table 3, the shortfall of pre-reversal I/B/E/S earnings per share relative to I/B/E/S analysts' forecasts is less than the difference between Compustat EPS and I/B/E/S forecasts. This difference is due to adjustments that I/B/E/S makes to reported ("actual") earnings. See footnote 8 for additional details.

TABLE 3
Actual Earnings and Earnings Benchmarks for Restructuring Charge Reversal Quarters and Non-Reversal Quarters

	<i>Reversal Quarters</i>		<i>Non-Reversal Quarters</i>		<i>p-value^c</i>
	<i>Mean</i>	<i>Median</i>	<i>Mean</i>	<i>Median</i>	
Actual earnings per share before reversal (from Compustat)	0.03	0.06	0.20	0.12	0.016
I/B/E/S earnings per share estimate ^a (median forecast of all analysts for each firm-quarter)	0.20	0.16	0.21	0.18	0.159
Shortfall of pre-reversal I/B/E/S earnings per share relative to I/B/E/S analysts forecasts ^b	0.09	0.02	0.03	0.00	0.001
Pre-reversal earnings per share decline from the same quarter of the prior year ^d	0.03	0.02	(0.07)	0.00	0.082

^a Sample size for analysts' forecast comparisons: $n = 1,831$ (83 reversal quarters and 1,748 non-reversal quarters).

^b This line represents the mean and median differences between the I/B/E/S median analysts' forecast for each firm-quarter observation and the I/B/E/S actual pre-reversal earnings for each firm-quarter observation.

^c p -values are one-tailed nonparametric analysis of variance tests.

^d Sample size for comparison to same quarter of prior year: $n = 2,555$ (121 reversal quarters and 2,434 non-reversal quarters).

to same-quarter-of-the-prior-year EPS in non-reversal quarters, but this difference is only marginally significant ($p = 0.08$ one-tailed test).

Frequency Analysis: Incidence of Reversals in Relation to Missed (Pre-Reversal) Analysts' Forecasts, Earnings Declines, and Losses

Table 4 provides evidence that managers are likely to record reversals in quarters when pre-reversal earnings fell short of analysts' forecasts, or same-quarter-prior-year earnings, or when pre-reversal earnings would be a net loss. Panel A of Table 4 shows that pre-reversal earnings fall short of analysts' forecasts in 71.1 percent of the 83 reversal quarters and only 37.8 percent of the 1,748 non-reversal quarters. Thus, managers are more likely

TABLE 4
Frequency Analysis: Incidence of Reversals in Relation to Missed (Pre-Reversal) Analysts' Forecasts, Earnings Declines, and Losses

	Reversal Quarters		Non-Reversal Quarters		Chi-Square p-value ^a
	Number	Percentage	Number	Percentage	
<i>Panel A: Reversals to Avoid Earnings Shortfall Relative to Analysts' Forecasts (n = 1,831)</i>					
• Number of firm-quarters examined	83	100.0	1,748	100.0	
• Number and percentage with pre-reversal earnings shortfall	59	71.1	661	37.8	0.001
• Number and percentage of pre-reversal shortfalls overcome by reversal	29	49.2	NA	NA	
• Number and percentage of quarters with post-reversal shortfall	30	36.1	661	37.8	0.380
<i>Panel B: Reversals to Avoid Earnings Decline from Same Quarter of Prior Year (n = 2,555)</i>					
• Number of firm-quarters examined	121	100.0	2,434	100.0	
• Number and percentage of quarters in which a pre-reversal earnings decline occurs	67	55.4	1,195	49.1	0.089
• Number and percentage of pre-reversal earnings declines that were overcome by reversal	18	26.9	NA	NA	
• Number and percentage of quarters with post-reversal earnings decline	49	40.5	1,195	49.1	0.065
<i>Panel C: Reversals to Avoid Reporting a Net Loss (n = 2,555)</i>					
• Number of firm-quarters examined	121	100.0	2,434	100.0	
• Number and percentage of quarters with a pre-reversal net loss	44	36.4	687	28.2	0.027
• Number and percentage of pre-reversal net losses negated by reversal	16	36.4	NA	NA	
• Number and percentage of quarters with a post-reversal net loss	28	23.1	687	28.2	0.224

^a One-tailed test.

to record reversals when pre-reversal earnings fall short of analysts' forecasts ($p = 0.001$, one-tailed test), supporting H1a. In 49.2 percent of the 59 reversal quarters in which pre-reversal earnings fell short of analysts' forecasts, the reversal was large enough to increase earnings to equal or exceed analysts' forecasts. As a result, shortfalls are observed in 36.1 percent of reversal quarters, which is not significantly different from the shortfalls in 37.8 percent of non-reversal quarters.

Panel B of Table 4 shows that pre-reversal earnings declines relative to the same quarter of the prior year occur in 55.4 percent of the 121 reversal quarters and in 49.1 percent of the 2,434 non-reversal quarters. These frequencies are marginally different ($p < 0.09$, one-tailed test), so the results marginally support H1b. The reversal was large enough to increase earnings to meet or exceed prior-year levels in 26.9 percent of the 67 pre-reversal decline quarters. Post-reversal declines are observed in 40.5 percent of reversal quarters, which is marginally significantly less ($p < 0.07$) than the 49.1 percent of non-reversal quarters that experience a decline.

I find pre-reversal net losses in 36.4 percent of the 121 reversal quarters, but only 28.2 percent of the 2,434 non-reversal quarters. Pre-reversal losses are significantly more common in reversal quarters than in non-reversal quarters, supporting H1c ($p < 0.03$). Furthermore, the reversal lifted earnings to zero or positive in 36.4 percent of the 44 pre-reversal loss quarters. As a result, losses are observed in 23.1 percent of reversal quarters, which is not significantly different from the 28.2 percent observed in non-reversal quarters.

Pre-reversal earnings fall short of the three benchmarks in reversal quarters more frequently than in non-reversal quarters. The reversals are sufficient that the proportion of reversal quarters with post-reversal shortfalls is less than (Panel B) or not significantly different from (Panels A and C) the proportion of non-reversal quarters with such shortfalls. Overall, 51 (42.1 percent) of the 121 reversals moved earnings from below to equal or above at least one of the three benchmarks.¹¹ In sum, the evidence presented in Table 4 suggests that reversals are more likely when pre-reversal earnings fall short of analysts' forecasts and when pre-reversal earnings are a net loss, supporting H1a and H1c. The results marginally support H1b, that reversals are more likely when pre-reversal earnings fall short of same-quarter-of-the-prior-year earnings.

OLS Regression Analysis: Magnitude of Reversals in Relation to Magnitude of Pre-Reversal Shortfalls from Analysts' Forecasts, Prior-Year Earnings, and Positive Net Income

The results from the OLS regression analysis of the magnitude of the restructuring charge reversal on the magnitude of pre-reversal shortfall from earnings benchmarks appear in Table 5. The significant positive analysts' forecast error coefficient (column 1) ($p = 0.0008$, one-tailed test) suggests that the amount of restructuring charge reversals that managers record increases with the magnitude of the pre-reversal earnings shortfall relative to analysts' forecasts. This evidence supports H2a. However, the magnitude of the earnings shortfall explains only a small proportion of the cross-sectional variation in the magnitude of the reversals.

The coefficient estimate associated with the amount by which prior-year earnings exceed pre-reversal earnings (column 2) is not significant. This evidence does not support H2b. However, the pre-reversal net loss coefficient (column 3) is significantly positive ($p < 0.03$, one-tailed test). This suggests that, on average, the larger the pre-reversal net

¹¹ In addition, 11 of these reversals caused earnings to move from below to above two or more earnings benchmarks.

TABLE 5
OLS Regression Analysis: Magnitude of Restructuring Charge Reversals in Relation to
Magnitude of Pre-Reversal Shortfalls from Analysts' Forecasts, Prior-Year Earnings, and
Positive Net Income

Analysts' Forecasts: Relation between the magnitude of the reversal and pre-reversal earnings shortfalls relative to analysts' forecasts:

$$REVPS_{ij} = \alpha_1 + \beta_1 SHORTFALLAF_{ij} + \varepsilon$$

Same Quarter of the Prior Year: Relation between the magnitude of the reversal and pre-reversal earnings declines from earnings in the same quarter of the prior year:

$$REVPS_{ij} = \alpha_2 + \beta_2 DECLINE_{ij} + \varepsilon'$$

Net Loss: Relation between the magnitude of the reversal and pre-reversal net losses:

$$REVPS_{ij} = \alpha_3 + \beta_3 (LOSSQTR \times |LOSSBR|)_{ij} + \varepsilon''$$

	<i>Analysts' Forecasts</i>	<i>Same Quarter of the Prior Year</i>	<i>Net Loss</i>
Intercept	0.0071	0.0129	0.0115
t-statistic	[4.765]	[5.958]	[5.064]
Significance (p-value) ^a	(0.0001)	(0.0001)	(0.0001)
Parameter estimate	0.0179	0.0001	0.0065
t-statistic	[3.164]	[0.133]	[1.955]
Significance (p-value) ^a	(0.0008)	(0.5528)	(0.0254)
Adjusted R ²	0.5%	0.04%	0.2%
Number of firm-quarters	1,831	2,555	2,555

^a p-values for estimated regression coefficients are based on t-statistics determined using heteroskedasticity-consistent standard error estimates (White 1980).

All amounts are adjusted to year-2000 dollars to control for inflation. Adjustment factors are based on the consumer price index.

REVPS (reversal amount per share) = the amount of the reversal firm *i* recorded in quarter *j*, divided by the number of common shares outstanding at the end of quarter *j*; SHORTFALLAF (earnings shortfall relative to analysts' forecasts) = the last median consensus I/B/E/S analysts' EPS forecast for firm *i* in quarter *j* minus pre-reversal actual EPS. I define pre-reversal actual earnings as (post-reversal actual earnings reported by I/B/E/S – reversal net of 40 percent tax). If the analyst forecast exceeds pre-reversal EPS, then SHORTFALLAF is positive; DECLINE (earnings decline from prior-year earnings) = actual EPS for firm *i* in quarter *j* of the prior year (excluding restructuring charges if any) minus actual pre-reversal EPS for firm *i* in quarter *j*. If the prior-year earnings exceed the current year pre-reversal earnings, then DECLINE is positive; LOSSQTR (loss quarter) = 1 if firm *i* has a pre-reversal net loss in quarter *j*, and 0 otherwise; and LOSSBR (loss per common share before reversal) = the absolute value of the actual pre-reversal loss per share for firm *i* in quarter *j*.

loss, the larger the reversal, supporting H2c. Again, however, the overall explanatory power of the magnitude of the loss is low.

The results in Tables 4 and 5 indicate that the incidence and magnitude of restructuring charge reversals increase as pre-reversal earnings fall short of earnings target benchmarks. In fact, 51 of the 121 reversals lifted earnings from below to equal or above at least one of the three earnings targets. However, 70 of the 121 reversals did not lift earnings to any of the three targets and the regressions explaining the magnitude of reversals have low

explanatory power. There are at least two likely explanations. First, many reversals likely reflect updates of earlier best-guess estimates, rather than earnings management, *per se*. Second, even to the extent that managers record reversals to manage earnings, this study examines only management toward three specific earnings benchmarks. Managers may face other incentives to manage earnings, such as those stemming from debt contracts. Detecting earnings management for contracting reasons requires contractual details, such as specific debt constraints, that are not publicly available. Future research might jointly model the earnings targets examined in this study and contract-based incentives, if the researcher can identify necessary contractual details.

Sensitivity Analyses

The regression inferences reported above hold after including a proxy for debt contracting incentives—the difference between the sample firm's debt-to-assets ratio and the weighted-average debt-to-assets ratio in the sample firm's (four-digit SIC code) industry.¹² I found this factor to be insignificant and including it in the analysis did not affect the study's inferences.

The primary analyses exclude firm-quarter observations that were at least \$5 below the benchmark, because a reversal would likely not be sufficient to cover this shortfall. Similar inferences arose after winsorizing these observations to \$5 shortfalls, using \$2 or \$1 shortfall thresholds, and without omitting these extreme observations. Regressions based on the ranks of dependent and independent variables also yielded similar inferences and overall explanatory power.

The models reported above use values scaled by the number of shares. As a sensitivity check, I estimated the models using total sales as an alternative scalar and using unscaled values. I also estimated models controlling for firm effects and time effects. Each of these analyses yields inferences consistent with those drawn above. The regression inferences also hold after I include shortfalls from all three benchmarks as separate independent variables in the same regression. Specifically, the coefficients on the pre-reversal shortfall relative to analysts' forecasts and on the amount of the loss are significantly positive ($p < 0.03$), but the coefficient on the shortfall relative to same-quarter-of-the-prior-year earnings is not significant.

VI. SUMMARY

Managers have incentives to use accounting discretion to manage earnings to meet or beat several targets, including analysts' forecasts, prior earnings levels, and avoiding losses. One source of accounting discretion available to managers is the accounting estimates associated with the financial reporting of restructuring activities. Regulators have expressed concern that firms use restructuring charge accruals to manage earnings. During the 1990s, increasing numbers of firms reversed portions of earlier restructuring charges.

In this study, I find evidence consistent with managers using restructuring charge reversals to manage earnings. Managers are more likely to record reversals when pre-reversal earnings are below analysts' forecasts or when the firm experiences a pre-reversal loss, and they are marginally more likely to record reversals in quarters when pre-reversal earnings are below prior-year levels. The amount of reversals increases with the amount by which earnings fall short of analysts' forecasts and the amount of pre-reversal net losses, although

¹² I also controlled for the existence of bonus-based incentive compensation contracts using an indicator variable equal to 1 if the CEO has a bonus plan based on accounting earnings, or 0 otherwise. Including this contracting incentive proxy does not affect the inferences from the regression.

these shortfalls explain only a modest portion of the cross-sectional variation in the magnitude of restructuring charge reversals. Consistent with regulators' and analysts' concerns, these findings suggest that firms, in some instances, use restructuring charge reversals to manage earnings to meet or beat earnings benchmarks. Moreover, the increasing incidence of restructuring charge reversals, and the evidence that managers are more likely to record reversals when pre-reversal earnings miss benchmarks, suggest that initial restructuring charges may include some hidden reserves.

The study is limited in that it examines only those reversals that firms disclose. Firms likely record, but do not separately disclose, smaller magnitude reversals. In addition, the modest explanatory power of the models of the magnitude of the reversals suggests that management toward these earnings targets is just one of many possible reasons for reversals. Future research into the role of contracting incentives on the reversal amount could provide additional information about the timing and amount of restructuring charge reversals.

REFERENCES

- Abarbanell, J., and R. Lehavy. 2000. Differences in commercial database reported earnings: Implications for inferences concerning analyst forecast rationality, the association between prices and earnings, and firm reporting discretion. Working paper, The University of North Carolina at Chapel Hill and University of California, Berkeley.
- Atiase, R. K., D. E. Platt, and S. Tse. 2001. Operational restructuring charges and post-restructuring performance: The potential effects of write-off amounts on measured performance. Working paper, The University of Texas at Austin.
- Barth, M. E., J. A. Elliott, and M. W. Finn. 1999. Market rewards associated with patterns of increasing earnings. *Journal of Accounting Research* 37 (2): 387–413.
- Bartov, E., D. Givoly, and C. Hayn. 2001. The rewards to meeting or beating earnings expectations. Working paper, New York University; University of California, Irvine; and University of California, Los Angeles.
- Beneish, M. D., and E. Press. 1993. Costs of technical violation of accounting-based debt covenants. *The Accounting Review* 68 (2): 233–257.
- Bens, D. A. 2000. Earnings management in the context of corporate restructurings. Working paper, University of Chicago.
- Brickley, J. A., and L. D. Van Drunen. 1990. Internal corporate restructuring: An empirical analysis. *Journal of Accounting and Economics* 12 (1–3): 251–280.
- Brown, L. D. 2001. A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research* 39 (2): 221–241.
- Buffett, W. E. 1999. Berkshire Hathaway 1998 letter to shareholders. Omaha, NE. March 1: 15. Available at: <http://www.berkshirehathaway.com/letters/1998pdf.pdf>.
- Burgstahler, D. C. 1997. Incentives to manage earnings to avoid earnings decreases and losses: Evidence from quarterly earnings. Working paper, University of Washington, Seattle.
- , and I. D. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24 (1): 99–126.
- , and M. Eames. 2001. Management of earnings and analyst forecasts to achieve zero earnings surprise. Working paper, University of Washington and Santa Clara University.
- Byrnes, N., and R. A. Melcher. 1998. Earnings hocus-pocus: How companies come up with numbers they want. *Business Week* (October 5): 134.
- Daley, L. A., and R. L. Vigeland. 1983. The effects of debt covenants and political costs on the choice of accounting methods: The case of accounting for R&D costs. *Journal of Accounting and Economics* 5 (3): 195–211.
- DeAngelo, H., L. DeAngelo, and D. Skinner. 1996. Reversal of fortune: Dividend signaling and the disappearance of sustained earnings growth. *Journal of Financial Economics* 40 (3): 341–371.
- Dechow, P. M., M. R. Huson, and R. G. Sloan. 1994. The effect of restructuring charges on executives' cash compensation. *The Accounting Review* 69 (1): 138–156.

- , R. G. Sloan, and A. P. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193–225.
- , and D. J. Skinner. 2000. Earnings management: Reconciling the views of accounting academics, practitioners, and regulators. *Accounting Horizons* 14 (2): 235–250.
- DeFond, M. L., and J. Jiambalvo. 1994. Debt covenant violation and manipulation of accruals. *Journal of Accounting and Economics* 17 (1–2): 145–176.
- DeGeorge, F., J. Patel, and R. Zeckhauser. 1999. Earnings management to exceed thresholds. *Journal of Business* 72: 1–33.
- Dichev, I. D., and D. J. Skinner. 2001. Large sample evidence on the debt covenant hypothesis. Working paper, University of Michigan.
- Elgers, P. T., R. J. Pfeiffer, Jr., and S. L. Porter. 2001. Anticipatory income smoothing: A re-examination. Working paper, University of Massachusetts Amherst.
- Elliott, J. A., and W. H. Shaw. 1988. Write-offs as accounting procedures to manage perceptions. *Journal of Accounting Research* 26 (Supplement): 91–119.
- , and J. D. Hanna. 1996. Repeated accounting write-offs and the information content of earnings. *Journal of Accounting Research* 34 (Supplement): 135–155.
- Emerging Issues Task Force (EITF). 1994. *Liability Recognition for Costs to Exit an Activity (Including Certain Costs Incurred in a Restructuring)*. Abstracts Issue No. 94-3. New York, NY: AICPA.
- Gaver, J. J., and K. M. Gaver, and J. Austin. 1995. Additional evidence of bonus plans and income management. *Journal of Accounting and Economics* 19: 3–28.
- , and ———. 1998. The relation between nonrecurring accounting transactions and CEO cash compensation. *The Accounting Review* 73 (2): 235–253.
- Gill, S., R. Gore, and L. Rees. 1996. An investigation of asset write-downs and concurrent abnormal accruals. *Journal of Accounting Research* 34 (Supplement): 157–169.
- Holthausen, R., D. F. Larcker, and R. G. Sloan. 1995. Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics* 19 (1): 29–74.
- Ip, G. 1996. Write-offs may become “write-ons.” *Wall Street Journal* (December 30): C1.
- Jennings, R., L. H. Martin, and R. B. Thompson II. 1998. The effects of corporate restructurings on the usefulness of the balance sheet. *Journal of Financial Statement Analysis* 3 (2): 44–56.
- John, K., L. Lang, and J. Netter. 1992. The voluntary restructuring of large firms in response to performance decline. *Journal of Finance* 47 (3): 891–917.
- Kasznik, R., and M. F. McNichols. 2001. Does meeting earnings expectations matter? Evidence from analyst forecast revisions and share prices. Working paper, Stanford University.
- Kinney, W., D. Burgstahler, and R. Martin. 2001. The materiality of earnings surprise. Working paper, The University of Texas at Austin, University of Washington, and Indiana University.
- Kross, W. J., T. Park, and B. T. Ro. 1996. The impact of operational restructuring announcements on stock price, risk and trading volume. Working paper, Purdue University.
- Levitt, A. 1998. The numbers game. Speech delivered at the NYU Center for Law and Business, New York, NY, September 28.
- Libby, R., and W. R. Kinney. 2000. Does mandated audit communication reduce opportunistic corrections to manage earnings to forecasts? *The Accounting Review* 75 (4): 383–404.
- Lopez, T. J., and L. Rees. 2001. The effect of meeting analysts’ forecasts and systematic positive forecast errors on the information content of unexpected earnings. Working paper, Texas A&M University.
- Matsunaga, S. R., and C. W. Park. 2001. The effect of missing a quarterly earnings benchmark on the CEO’s annual bonus. *The Accounting Review* 76 (3): 313–332.
- Richardson, S., S. H. Teoh, and P. Wysocki. 2001. The walkdown to beatable analyst forecasts: The roles of equity issuance and insider trading incentives. Working paper, University of Michigan—Ann Arbor, The Ohio State University, and Massachusetts Institute of Technology.
- Schuetz, W. 1998. *Update of Activities of the SEC Enforcement Division*. SEC Financial Reporting Conference, May, Pasadena, CA.

- Skinner, D. J., and R. G. Sloan. 2001. Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. Working paper, University of Michigan—Ann Arbor.
- Smith, R., and S. Lipin. 1996. Are companies using restructuring costs to fudge the figures? *Wall Street Journal* (January 30): A1.
- Sweeney, A. 1994. Debt covenant violations and managers' accounting responses. *Journal of Accounting and Economics* 17 (3): 281–308.
- Watts, R., and J. Zimmerman. 1986. *Positive Accounting Theory*. Englewood Cliffs, NJ: Prentice Hall.
- , and ———. 1990. Positive accounting theory: A ten-year perspective. *The Accounting Review* 65 (1): 131–156.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4): 817–838.

International Diversification and Analysts' Forecast Accuracy and Bias

Augustine Duru

David M. Reeb

American University

ABSTRACT: We investigate the association between corporate international diversification and the accuracy and bias of consensus analysts' earnings forecasts. We find that greater corporate international diversification is associated with less accurate and more optimistic forecasts. Our results suggest that international diversification reflects unique dimensions of forecasting difficulty that are not captured in previously identified determinants. This evidence suggests that as firms become more geographically diversified, forecasting their earnings becomes more complex.

Keywords: *analysts' earnings forecasts; corporate diversification; internationalization; forecast accuracy; forecast bias.*

Data Availability: *The data are publicly available from the sources indicated.*

I. INTRODUCTION

Over the past decade, U.S. firms have substantially expanded their operations abroad. Cross-border investments, for example, have grown by more than 700 percent (World Trade Organization 2000), and S&P 500 firms report that foreign sales, on average, account for more than 24 percent of total sales. Although prior literature shows that geographic segment disclosures help analysts develop more accurate earnings forecasts (Kinney 1971; Collins 1976; Baldwin 1984; Roberts 1989; Balakrishnan et al. 1990; Nichols et al. 1995), there is surprisingly little evidence on how corporate international diversification affects the properties of analysts' forecasts. This study investigates how corporate international diversification affects the accuracy and bias of analysts' earnings forecasts.

We are grateful for helpful comments from an *ad hoc* associate editor and two anonymous reviewers. We also thank Ajay Adhikari, Ron Anderson, Sudipta Basu, Huong Higgins, Karen Olender, Leigh Riddick, Michel Robe, Bob Thompson, Segun Wallace, Jingsan Zhu, and seminar participants at American University and the 2000 AAA Annual Meeting for their comments and suggestions. We also acknowledge research support received as Kogod Endowed Faculty Fellows.

*Submitted October 1999
Accepted November 2001*

We argue that corporate international diversification increases the complexity of the forecasting task. Because the accuracy of analysts' earnings forecasts declines as the complexity of the forecasting task increases (e.g., Brown, Richardson, and Schwager 1987; Bhushan 1989; Haw et al. 1994; Lang and Lundholm 1996; Clement 1999), we expect corporate international diversification to be associated with lower forecast accuracy.

We also examine the relation between international diversification and the directional bias in analysts' forecasts. Lim (2001) provides theory and evidence that analysts' earnings forecasts are more optimistically biased for firms with less predictable earnings. Similarly, Das et al. (1998) report a positive relation between optimistic bias and the unpredictability of analysts' earnings forecasts. Thus, to the extent that international diversification is associated with less predictable earnings, we also expect it to be associated with more optimistic analysts' forecasts.

Our evidence supports the prediction that international diversification is associated with less accurate and more optimistic earnings' forecasts, even after we control for previously identified determinants of forecasting difficulty (e.g., the variability in earnings, forecast horizon, firm size, and industrial diversification). This evidence suggests that international diversification reflects a unique dimension of forecasting complexity or difficulty that is incremental to the factors identified in prior research.¹

Accounting researchers, investors, and analysts are interested in obtaining more accurate earnings forecasts as a proxy for the capital markets' expectation of earnings, and for use in firm valuation (Schipper 1991). Identification of an additional, important determinant of forecast accuracy and bias can be valuable to these users of accounting information. For example, our findings indicate that researchers should control for international diversification in studies examining cross-sectional differences in the relation between price and expected future earnings. Because analysts' forecasts are, on average, less accurate for international firms, valuation of these firms is more difficult, suggesting that such studies should control for international diversification. Our results also suggest that when investors and researchers use analysts' forecasts as a basis for valuing internationally diversified firms, they should adjust the forecasts for the optimistic bias. More broadly, our evidence increases our understanding of the cross-sectional factors that influence the accuracy and bias of analysts' earnings forecasts.

Our evidence is also relevant to policymakers. The increasing importance of foreign operations in multinational firms has spurred the accounting and financial communities to improve the reliability of financial information about multinational firms by globalizing financial intermediaries and introducing new reporting standards, including the Statement of Financial Accounting Standards (SFAS) No. 131. SFAS No. 131 (FASB 1997), which replaced SFAS No. 14 (FASB 1976), improves business segment reporting; however, the disclosure of earnings by geographic area has declined significantly (Herrmann and Thomas 2000) because SFAS No. 131 does not require the disclosure of foreign income in enterprise-wide disclosures. Our evidence that international diversification is associated with less accurate and more optimistic analysts' earnings forecasts suggests that market participants may benefit from expanded reporting requirements for international operations that provide more information on multinational firms' risks and opportunities.

The paper is organized as follows. In Section II, we discuss the related literature and present our hypotheses. Section III describes our sample, data, and variable measures. In

¹ In related research, Das and Saudagaran (1998), Basu et al. (1998), and Ashbaugh and Pincus (2001) investigate the properties of analysts' forecasts for non-U.S. firms and find that forecast accuracy depends on the information environment. In contrast to our study, this prior research does not consider either U.S. multinational firms or the effect of international diversification.

Section IV, we describe the research design and present the empirical results. In Section V, we test the robustness of our results by using alternative measures and specifications. Section VI summarizes and concludes.

II. HYPOTHESIS DEVELOPMENT

International Diversification and Forecast Accuracy

Several studies have examined the determinants of forecast accuracy (e.g., Brown, Richardson, and Schwager 1987; Kross et al. 1990; Haw et al. 1994; Wiedman 1996; Lang and Lundholm 1996; Clement 1999). Brown et al. (1985) and Brown (1993) review this literature and conclude that the accuracy of an earnings forecast depends on the difficulty or complexity of the forecasting task. Although firms report geographic segment information under SFAS No. 14 (or SFAS No. 131), we do not expect these disclosures to fully offset the greater complexity of predicting foreign earnings.² Specifically, we argue that international diversification reflects unique dimensions of the forecasting task and of environmental complexity not captured in previously identified proxies.

One source of the greater complexity of international operations stems from the greater volatility of foreign earnings relative to that of domestic earnings. Although earlier studies (e.g., Shapiro 1978) argue that international diversification decreases earnings volatility due to a portfolio effect (i.e., uncorrelated cash flows in multiple countries), subsequent studies find that corporate international diversification is also associated with an environmental effect that increases earnings volatility. For example, Goldberg and Heflin (1995) and Reeb et al. (1998) find that corporate geographic diversification increases firms' exposure to economic factors (e.g., currency risk and political risk), regulatory intervention, and turbulence, which increase earnings volatility.³ In our sample, the environmental effect dominates the portfolio effect, resulting in an increase in earnings volatility: the foreign income of U.S. multinationals has about a 65 percent greater coefficient of variation than does their domestic income. In sum, at least in our sample, international diversification leads to earnings volatility and consequently we hypothesize less accurate earnings forecasts.

A second source of forecasting difficulty arises because analysts typically have more knowledge about their home countries than they do of foreign countries, suggesting that they are likely to have less information about a firm's foreign operations than about its domestic operations (Ashbaugh and Pincus 2001). Different cultures, customs, competitors, geographic features, and regulations imply that analysts have less-specialized knowledge of foreign markets that they can use to interpret, analyze, and understand a firm's competitive positions. These cultural differences, geographic constraints, differing legal systems, and language differences increase the difficulty of forecasting international operations.

Managers also enjoy more discretion as a firm increases operations abroad, which in turn increases information asymmetry between analysts and management. Kogut (1983) argues that geographic diversity increases multinational firms' operational flexibility and allows firms to increase value by exploiting the increased uncertainty of the international environment. For example, global manufacturing gives managers additional opportunities to exercise discretion by shifting production to lower-cost locations. Bodnar et al. (1999) note that operating in multiple geographic locations creates additional options, including

² SFAS No. 14 (FASB 1976), *Financial Reporting for Segments of a Business Enterprise*, requires firms to disclose geographic segment information on foreign operations whose sales or assets are greater than 10 percent of consolidated sales or assets. SFAS No. 131 (FASB 1997), *Disclosures about Segments of an Enterprise and Related Information*, which became effective for firms with year-ends after December 15, 1998, requires a management approach in identifying segments. SFAS No. 131 defines a segment as a component of the enterprise for which separate financial information is available and is reviewed regularly by management.

³ For example, Gay and Nam (1998) argue that most firms do not completely hedge exchange rate risk.

the ability to arbitrage temporary international market imperfections and to transfer profits or losses within the firm to take advantage of international tax differences. To the extent that arbitrage opportunities are not readily observable, analysts will find it difficult to predict their exercise, and thus will make less accurate forecasts.

Alternatively, more information could be available about multinational firms because they are usually larger than purely domestic firms. However, we argue that analysts' lower familiarity with foreign operating environments and the greater information asymmetry between analysts and managers outweighs the effects of the increased public and private information. In addition, although recent research suggests that foreign earnings are more persistent than domestic earnings (Bodnar and Weintrop 1997; Thomas 2000), income from non-U.S. operations could still be more difficult to predict if analysts' unfamiliarity with the international operations prevents them from appreciating this persistence. This argument is consistent with Thomas's (2000) evidence that investors fail to appreciate fully the greater persistence of income from non-U.S. operations.

The preceding discussion suggests that international diversification is associated with unique dimensions of forecasting difficulty that are incremental to the proxies used in prior studies. Managers' additional opportunities to exercise discretion can increase earnings volatility, and the difficulty of predicting future earnings. Analysts' unfamiliarity with the foreign business environment suggests that analysts will have difficulty processing data on foreign operations into accurate forecasts. Therefore, we predict that firm internationalization is incrementally useful in explaining forecast accuracy, even after controlling for previously identified determinants of forecast accuracy:

- H1:** Greater international diversification is associated with less accurate mean consensus analysts' earnings forecasts, even after controlling for previously identified determinants of forecast accuracy.

International Diversification and Optimistic Bias

Previous studies also investigate determinants of cross-sectional differences in analysts' forecast optimism. Under the premise that management prefers favorable forecasts, Lim's (2001) model of rational bias posits that analysts issue more favorable forecasts for firms with less predictable earnings in order to maintain favorable relations with managers who allow those analysts to obtain access to private information, and thereby improve the accuracy of the analysts' forecasts. Consistent with Lim's (2001) prediction, Das et al. (1998) provide empirical evidence that forecasts are more optimistically biased when earnings are less predictable.

Because we expect international operations to increase the difficulty of predicting income, we posit international diversification to be associated with more optimistic analysts' forecasts. We expect international diversification to have an incremental effect on analysts' optimism, after controlling for other determinants of optimistic bias, because of the increment in the unpredictability of income due to additional opportunities for managerial discretion and to analysts' unfamiliarity with the international information environment.⁴

⁴ Prior research has explored several other explanations for optimistic bias. For example, a stream of literature (Downen 1996; Hwang et al. 1996; Degeorge et al. 1999; Brown 2001) argues that analysts' forecasts are more optimistically biased for loss firms than for profitable firms. McNichols and O'Brien (1997) posit that optimistic bias is the result of self-selection, whereby analysts who hold unfavorable views about the firm do not provide forecasts. Gu and Wu (2001) provide evidence to suggest that optimistic bias is induced by negatively skewed

(Continued on next page)

H2: Greater international diversification is associated with more optimistically biased mean consensus analysts' earnings forecasts, even after controlling for previously identified determinants of forecast bias.

Our sample period, 1995–1998, is before SEC's Regulation Fair Disclosure (FD), which became effective on October 23, 2000. Regulation FD affects the subsequent applicability of H2, because it prohibits firms from disclosing material new information to select securities market participants, including analysts. Hypothesis 2 is based on the premise that analysts bias forecasts upward to gain access to management's nonpublic information, but Regulation FD makes it less likely that management will provide analysts with nonpublic information in the future, regardless of the nature of the forecasts.

III. DATA DESCRIPTION AND VARIABLE MEASUREMENT

Sample Selection

Of the about 6,000 U.S. firms in the Disclosure Worldscope Database (DWS), 1,217 firms meet two data requirements. First, we require firms to report foreign assets and foreign sales. Second, we require that the Institutional Brokers Estimation System (I/B/E/S) annual consensus earnings forecast, which is provided in the database, be based on at least three analysts' forecasts. Of the 1,217 sample firms, 530 (approximately 43 percent) report zero foreign assets, zero foreign sales, and zero geographic segments, so we consider them purely domestic firms.

We measure consensus forecasts as either the mean or the median forecasts. The DWS database provides I/B/E/S consensus earnings forecasts for the current year and for the following year. We use forecasts for the fiscal year obtained from the March monthly update.⁵ Because most firms have December 31 fiscal year-end dates, most of our sample has a nine-month forecast horizon. In our sample, firms with April through November fiscal year-end dates have shorter forecast horizons, and firms with a January, February, or March fiscal year-end date (which have already reported annual earnings in the current calendar year) have a longer forecast horizon because we use the forecast for the following calendar year. Thus, the forecast horizons in the sample range from 1 to 12 months. These data requirements result in a sample of 3,495 firm-year observations for the period 1995–1998.

Measuring International Diversification

Sullivan (1994) notes that the most common proxy for corporate international diversification is the foreign sales ratio (i.e., foreign sales divided by total sales). However, this sales-based proxy mixes export sales with foreign subsidiary sales and ignores geographic structural differences among firms, such as asset placement. Reeb et al. (1998) note that

Footnote 4 (Continued)

earnings. In addition, Brown (1997, 2001) shows that optimistic bias has decreased over time and that it is negatively associated with firm size and analyst following. Our empirical tests control for all of these determinants. However, there are other alternative incentives for optimistic bias. For example, Hayes (1998) argues that self-selection arises because analysts focus on firms they expect to perform well in order to generate higher trading commissions for their brokerage firms. Similarly, others conclude that analysts are more optimistic in the presence of underwriting opportunities (e.g., Dugar and Nathan 1995; Lin and McNichols 1998; Dechow et al. 2000). These latter arguments focus on systematic differences in the properties of analysts' forecasts between groups of analysts. In contrast, we concentrate on consensus analysts' earnings forecasts.

⁵ We use the March monthly update because of data availability in the DWS. Specifically, because the DWS database provides I/B/E/S consensus earnings forecasts for the current year and for the following year (and does not provide historical forecasts), we obtained updates of the consensus forecast of annual earnings from the database's four annual CD-ROMs. Robustness tests reported in Section V suggest that differences in forecast horizon cannot explain our results.

the foreign asset ratio (i.e., foreign assets divided by total assets) mitigates the problem of mixing export and foreign subsidiary sales, and it captures geographic structural information. However, the foreign asset ratio does not fully capture the complexity of the forecasting task, because it does not capture the number of different markets in which the firm operates.

Therefore, we construct a composite measure of the degree of international diversification (ID), using factor analysis to reduce three common measures of international diversification into a single index: the foreign sales ratio, the foreign asset ratio, and the number of geographic segments in which the firm operates. During our sample period, firms reported geographic segment data in accordance with SFAS No. 14, which required disclosure of geographic segments comprising more than 10 percent of the firm (e.g., by sales).⁶

Factor analysis requires that attribute measures be highly intercorrelated and load on a single factor. Table 1, Panel A shows that the correlations among the foreign sales ratio, the foreign asset ratio, and the number of geographic segments exceed 65 percent. Panel

⁶ SFAS No. 131, which replaced SFAS No. 14, became effective for firms with year-ends after December 15, 1998.

TABLE 1
Measuring International Diversification in a Sample of 1,217 Firms (3,495 Firm-Year Observations) for the Period 1995–1998
Factor Analysis Results^a

Panel A: Pearson Correlation Matrix

	<u>Foreign Asset Ratio</u>	<u>Foreign Sales Ratio</u>	<u>Number of Geographic Segments</u>
Foreign Asset Ratio	1.00		
Foreign Sales Ratio	0.76	1.00	
Number of Geographic Segments	0.66	0.74	1.00

Panel B: Total Variance Explained^b

<u>Component</u>	<u>Eigenvalue</u>	<u>% of Variance</u>	<u>Cumulative %</u>
1	2.44	81.16	81.16
2	0.34	11.41	92.56
3	0.22	7.44	100.00

Number of Panel C: Component Matrix/Factor Loadings^c

<u>Component</u>	<u>Loading</u>
Foreign Asset Ratio	0.89
Foreign Sales Ratio	0.92
Number of Geographic Segments	0.88

^a Results from computing the degree of international diversification (ID) measure using factor analysis based on three attributes of international activity (foreign asset ratio, foreign sales ratio, and the number of geographic segments in which a firm operates).

^b Component extracted: Only component one has an eigenvalue greater than 1.

^c Extraction method: Principal component analysis.

B of Table 1 indicates that these three attributes load on a single factor that explains 81 percent of the cumulative variance, and Panel C reports the factor loadings. The results in Table 1 suggest that the factor analysis provides a useful index of internationalization. Untabulated results show that using the foreign sales ratio instead of our index yields identical inferences.

Measuring Forecast Accuracy (ACCURACY) and Bias (BIAS)

We measure forecast accuracy for each firm-year observation based on the absolute value of the difference between the firm's forecasted and actual earnings, divided by the stock price at the forecast date:

$$\text{ACCURACY}_t = (-1) \frac{|\text{FORECAST}_t^{t-1} - \text{EARN}_t|}{\text{PRICE}_{t-1}} \quad (1)$$

where ACCURACY_t is the negative of the consensus absolute forecast error at time t , FORECAST_t^{t-1} is the consensus forecast of period t earnings made at period $t - 1$, EARN_t is the actual earnings per share before extraordinary items for period t (from the DWS, which is also identical to the Compustat reported earnings), and PRICE_{t-1} is the stock price at the time of the forecast ($t - 1$). Multiplying this absolute forecast error by (-1) gives a measure that increases with greater forecast accuracy. Thus, a firm attribute that is negatively associated with ACCURACY signals that the attribute indicates less accurate analysts' forecasts.

As in previous research, we measure bias (BIAS) as the signed forecast error, which we define as the difference between the consensus earnings forecast and the actual earnings, scaled by the stock price:

$$\text{BIAS}_t (\text{OPTIMISM}) = \frac{\text{FORECAST}_t^{t-1} - \text{EARN}_t}{\text{PRICE}_{t-1}} \quad (2)$$

Forecast optimism increases as BIAS becomes larger. Thus, a firm attribute that is positively associated with BIAS indicates that the attribute is associated with more optimistic analysts' forecasts. Our results are robust when we use the absolute value of the mean of the consensus forecast as the deflator for ACCURACY and for BIAS .

Controlling for Other Determinants of Forecast Accuracy and Bias

Our multiple regressions control for previously identified determinants of forecast accuracy and bias: task complexity attributes, earnings attributes, and forecast attributes.

Other Factors that Affect Task Complexity

Factors other than internationalization that increase task complexity include industrial diversification (IND) and firm size (SIZE). Dunn and Nathan (2000) suggest that industrial diversification is associated with less accurate and more biased (optimistic) analysts' forecasts. We measure industrial diversification using two-digit SIC codes. Our measure incorporates both the number of industries in which a firm operates and the relative magnitude of the industrial diversification:

$$\text{IND} = [\text{UNREL}/\text{SALES}] \times \text{NUM} \quad (3)$$

where UNREL is the sum of the sales outside the firm's primary two-digit SIC code, SALES

represents the firm's total sales, and NUM is the number of two-digit SIC code industries (outside the primary SIC code) in which the firm reports sales.⁷

Firm size has two potential and opposing effects. Firm size likely indicates more complexity and thus greater forecast error, but more predisclosure information is available for larger firms, which might lead to lower forecast error. We measure firm size as the natural log of total assets.

Other Attributes of Earnings That Increase Forecasting Difficulty

Other attributes of earnings that increase forecasting difficulty include the reporting of net losses (LOSS), earnings skewness (SKEW), the magnitude of current changes in earnings ($\Delta EARN$), and longer-term earnings volatility (EVOLAT). Prior studies (e.g., Hwang et al. 1996) find that analysts' forecasts of losses are, on average, less accurate than forecasts of profits. Brown (2001) provides evidence that analysts, on average, issue more optimistic forecasts in loss periods. Managers of loss-reporting firms may also understate reported earnings to facilitate reporting higher profits in future periods (i.e., "big bath" incentives). To control for the effect of losses, we include an indicator variable that takes a value of 1 for loss firm-years, and 0 otherwise. To the extent that analysts expect losses, controlling for losses also helps control for selection bias (McNichols and O'Brien 1997). We expect loss firm-years to be associated with less accurate and more optimistic analysts' forecasts.

Previous research suggests that earnings skewness and the magnitude of the current change in earnings also affect analysts' forecasts (Brown, Richardson, and Schwager 1987; Lang and Lundholm 1996; Gu and Wu 2001). Gu and Wu (2001) provide empirical evidence that skewed earnings are associated with more optimistically biased forecasts. Lang and Lundholm (1996) find that larger changes in earnings are associated with less accurate forecasts. We control for earnings skewness (SKEW) based on the time-series of earnings over the past ten years, using the statistical definition of skewness.⁸ We measure the change in earnings ($\Delta EARN$) as the absolute value of the difference between the current year's earnings per share and last year's earnings per share, divided by the price at the beginning of the year. Prior research suggests that longer-term earnings volatility (EVOLAT) is associated with less accurate and more optimistically biased forecasts (Kross et al. 1990; Lim 2001). Thus, we also control for earnings volatility, measured as the standard deviation of return on assets for the previous five-year period.

Forecast Attributes

Forecast attributes include forecast horizon (HOR), the number of analysts following the firm (FOLLOW), and forecast dispersion (DISP).⁹ Previous research (e.g., Brown, Hagerman, Griffin, and Zmijewski 1987; Brown, Richardson, and Schwager 1987; O'Brien and Bhushan 1990; Brown 1993) indicates that longer forecast horizons are associated with less accurate analysts' earnings forecasts. Kang et al. (1994) and Das et al. (1998) find

⁷ Alternative measures of industrial diversification (e.g., three-digit SIC codes, the number of industries, and a Herfindal index) lead to similar results.

⁸ This is the commonly used skewness coefficient defined for firm j as:

$$SKEW_j = \frac{n_j}{(n_j - 1)(n_j - 2)} \sum_t ((x_{jt} - \bar{x}_j)/s_j)^3$$

where x_{jt} is actual earnings per share before extraordinary items for period t and firm j , $\bar{x}_j$ is the mean and s_j the standard deviation of x_{jt} (based on n_j observations for each firm).

⁹ Both the number of analysts following the firm and forecast dispersion are also related to task complexity. The more analysts following a firm, the more predisclosure information. In contrast, the dispersion of the forecasts is a contemporaneous measure of the complexity inherent in developing the forecasts.

weak evidence that longer horizon forecasts are more optimistic. We measure forecast horizon (HOR) in months using the DWS data (from I/B/E/S).

Lys and Soo (1995) find that, after controlling for firm-specific forecasting difficulty, forecast accuracy increases with analyst following, whereas Das et al. (1998) suggest less optimistic forecasts for more heavily followed firms. We control for analyst following (FOLLOW) using the number of analysts whose forecasts appear in the I/B/E/S system.

Previous studies (e.g., Lang and Lundholm 1996; Bamber et al. 1997) suggest that forecast dispersion proxies for uncertainty and lack of consensus among analysts about future events. Lang and Lundholm (1996) find that forecast dispersion is negatively related to analysts' forecast accuracy. Furthermore, Gu and Wu (2001) suggest that dispersion is associated with less optimistically biased forecasts. As in Lang and Lundholm (1996), we measure forecast dispersion (DISP) as the standard deviation of analysts' earnings forecasts, deflated by the stock price at the forecast date.

We also include dummy variables to control for industry affiliation (INDUSTRY), based on two-digit SIC codes, and calendar year (YEAR), because forecasting difficulty differs across industries and across time periods.

IV. EMPIRICAL ANALYSIS

Descriptive Statistics

Table 2, Panel A provides descriptive statistics for the key variables. The mean and median ACCURACY values are negative by construction. The closer the accuracy value is to zero, the more accurate the forecast. Thus, variables that are negatively related to ACCURACY are associated with less accurate forecasts.

The mean and median signed forecast errors (BIAS) are positive, indicating that, in our sample annual earnings forecasts for 1995 to 1998 (average nine-month horizon), analysts were, on average, optimistic, consistent with Easterwood and Nutt (1999), Abarbanell and Lehavy (2000), Gu and Wu (2001), and Richardson et al. (2001).¹⁰ Variables that are positively related to BIAS are associated with more optimistic forecasts. Mean and median consensus forecast data give similar results, so we report only the results based on the mean forecasts.

Table 2, Panel B shows the correlation matrix for the variables in our regression analysis (Spearman above the diagonal and Pearson below). We find that firm internationalization is associated with less accurate and more optimistic forecasts, consistent with our hypotheses. We also find that analysts issue less accurate and more optimistic forecasts in loss firm-years, as did Brown (2001).¹¹

¹⁰ Our finding that forecast biases are, on average, optimistic is consistent with evidence from recent studies that also use long-horizon forecasts of annual earnings from I/B/E/S. For example, Brown (1997) (with a sample of annual horizon forecasts covering from -2,109 to 4,593 firms between 1984 and 1996; see Brown [1997, footnote 12]) and Easterwood and Nutt (1999) (with a sample of eight-month horizon forecasts issued for 1,608 firms over the period 1982 to 1995) find an optimistic average bias for every year of their respective samples. Richardson et al. (2001) find similar results, using a subsample consisting of 28,246, 12-month horizon forecasts of annual earnings from 1983 to 1998 (although they find a pessimistic average bias in shorter horizon forecasts). In contrast, Brown (2001) and Matsumoto (2002) find median pessimistic bias. Matsumoto (2002) uses quarterly consensus forecasts with an average horizon of about three months from the Zack's database (Zack's drops forecasts that are older than 120 days). Brown (2001) uses I/B/E/S data, but focuses on the single (individual), most timely forecast based on quarterly forecasts. Thus, sample-selection procedures and the length of the forecast horizon can affect the size and sign of earnings surprise or bias, and this likely explains the differences in the studies' inferences regarding average optimism.

¹¹ Untabulated analyses indicate that loss firm-years play an important role in explaining forecast error and optimism. After we exclude loss firms from our sample, ACCURACY and BIAS have mean values closer to zero. We find that mean ACCURACY is -0.03 and mean BIAS is 0.02 for the profitable firms in our sample. Like Gu and Wu (2001), we find that profitable firms have some optimistic bias.

TABLE 2

Descriptive Analysis of Variables Related to Forecast Accuracy and Bias, for a Sample of 3,495 Firm-Year Observations for the Period 1995-1998

Panel A: Summary Statistics^a

	<u>Mean</u>	<u>Standard Deviation</u>	<u>Median</u>
Mean Consensus Forecast	2.15	1.74	1.82
Median Consensus Forecast	2.15	1.73	1.81
<i>Dependent Variables</i>			
Negative Absolute Forecast Error (ACCURACY)	-0.05	0.18	-0.02
Signed Forecast Error (BIAS)	0.04	0.19	0.02
<i>Hypothesized Variable</i>			
Degree of International Diversification (ID)	0.83	1.05	0.20
<i>Control Variables</i>			
Industry Diversification (IND)	0.81	0.67	0.80
Firm Size (SIZE)	10.08	0.99	10.40
Loss Firm-Years (LOSS)	0.11	0.31	0.00
Earnings Skewness (SKEW)	-0.03	0.09	0.02
Change in Earnings (Δ EARN)	0.00	0.18	-0.01
Earnings Volatility (EVOLAT)	0.14	0.30	0.06
Forecast Horizon (HOR)	8.19	2.41	9.00
Number of Analysts Following (FOLLOW)	11.71	7.94	9.00
Forecast Dispersion (DISP)	0.13	0.18	0.08

(Continued on next page)

TABLE 2 (Continued)

Panel B: Correlation Matrix^b

	ACCURACY	BIAS	ID	IND	SIZE	LOSS	SKEW	ΔEARN	EVOLAT	HOR	FOLLOW	DISP
ACCURACY												
BIAS	-0.96***		-0.22***	0.03	0.03	-0.30***	-0.42***	-0.87***	-0.16***	-0.06**	0.05**	-0.06**
ID	-0.22***	0.21**	0.24***	-0.03	0.03	0.31***	0.38***	0.87***	0.09**	0.06**	-0.04*	0.02
IND	0.04	-0.03	0.02	0.03	0.01	0.22***	0.11**	0.16**	0.09**	0.05**	0.18***	0.15**
SIZE	0.03	0.03	0.11**	0.04*	0.02	-0.06**	-0.05**	-0.03	-0.05**	0.16***	0.12**	0.08**
LOSS	-0.28***	0.30***	0.23***	-0.07**	0.02	0.02	0.28***	0.24***	0.17***	-0.01	-0.03	0.16**
SKEW	-0.42***	0.38***	0.11***	-0.07**	0.02	0.30***	0.18***	0.18***	0.17***	0.06**	-0.07*	-0.29***
ΔEARN	-0.87***	0.88***	0.15***	-0.02	0.03	0.25***	0.18***		0.02	0.03	-0.01	-0.03
EVOLAT	-0.16***	0.09**	0.10**	-0.05*	-0.02	0.15***	0.17***	0.04*		0.00	-0.04*	0.11**
HOR	-0.07**	0.06**	0.05*	0.16***	-0.03	-0.03	0.06**	0.03	0.01		0.11**	0.17***
FOLLOW	0.05**	-0.04*	0.20***	0.13***	0.73***	-0.03*	-0.07**	-0.01	-0.03	0.13**		0.14***
DISP	-0.05**	0.03	0.15***	0.07***	0.04	0.17***	-0.29***	-0.03	0.11**	0.16***	0.16***	

(Continued on next page)

TABLE 2 (Continued)

	Entire Sample		International Firms	
	Mean	Standard Deviation	Mean	Standard Deviation
Foreign Sales Ratio	0.12	0.19	0.21	0.21
Foreign Asset Ratio	0.13	0.20	0.23	0.22
Number of Geographic Segments	0.99	1.37	1.67	1.43

***, **, * Significant at the 1 percent, 5 percent, and 10 percent levels, respectively.

^a ACCURACY is the absolute forecast error multiplied by -1; BIAS is the signed forecast error, defined as the difference between the consensus earnings forecast and the actual earnings, scaled by the stock price; ID is international diversification, computed using factor analysis based on three attributes of international activity (foreign asset and sales ratios, and the number of geographic segments in which a firm operates); IND is industrial diversification, computed as the product of the unrelated sales ratio and the number of two-digit SIC code industries (outside the primary SIC code at the forecast date) in which the firm reports sales; SIZE is firm size, measured as the natural log of total assets; LOSS is coded as 1 for firm-year observations with negative earnings, and 0 otherwise; SKEW is earnings skewness, based on the prior ten years of earnings; ΔEARN is the change in earnings, measured as the absolute value of the difference between the current year's earnings per share and the previous year's earnings per share, divided by the price at the beginning of the year; EVOLAT is earnings volatility, measured as the standard deviation of return on assets for the previous 5-year period; HOR is the forecast horizon, expressed as the number of months between the forecast and the end of the fiscal year; FOLLOW is the number of analysts following the firm per I/B/E/S; DISP is forecast dispersion, measured using the standard deviation of analysts' forecasts deflated by the stock price at the forecast date.

^b Spearman coefficients appear above the diagonal line and Pearson coefficients below.

^c This panel provides information on the sample firms' international activity. The first two columns show data for the entire sample. The last two columns report this information for only those firms that report some international activity (i.e., ID > 0).

The first two columns of Panel C in Table 2 display the mean and standard deviation of three measures of international activity: foreign asset ratio, foreign sales ratio, and the number of geographic segments. We also present information for the subset of firms reporting international operations. Column 3 shows that the typical international firm in the sample has about 20 percent of its operations outside the U.S., so international activity is a material part of firms' operations.

Primary Tests

We use multiple regressions to test our hypotheses that, even after we control for previously identified determinants of forecast accuracy and bias, firm internationalization is associated with less accurate (H1) and more optimistic (H2) analysts' earnings forecasts:

$$\begin{aligned} \text{ACCURACY}_t = & A_0 + A_1(\text{ID}_{t-1}) + A_2(\text{IND}_{t-1}) + A_3(\text{SIZE}_{t-1}) + A_4(\text{LOSS}_t) \\ & + A_5(\text{SKEW}_{t-1}) + A_6(\Delta\text{EARN}_t) + A_7(\text{EVOLAT}_{t-1}) + A_8(\text{HOR}_{t-1}) \\ & + A_9(\text{FOLLOW}_{t-1}) + A_{10}(\text{DISP}_{t-1}) + A_{11}(\text{YEAR}_t) \\ & + A_{12}(\text{INDUSTRY}_{t-1}) + e \end{aligned} \quad (4)$$

$$\begin{aligned} \text{BIAS}_t = & B_0 + B_1(\text{ID}_{t-1}) + B_2(\text{IND}_{t-1}) + B_3(\text{SIZE}_{t-1}) + B_4(\text{LOSS}_t) + B_5(\text{SKEW}_{t-1}) \\ & + B_6(\Delta\text{EARN}_t) + B_7(\text{EVOLAT}_{t-1}) + B_8(\text{HOR}_{t-1}) + B_9(\text{FOLLOW}_{t-1}) \\ & + B_{10}(\text{DISP}_{t-1}) + B_{11}(\text{YEAR}_t) + B_{12}(\text{INDUSTRY}_{t-1}) + e' \end{aligned} \quad (5)$$

where ID is the multidimensional international diversification index, IND is industrial diversification, SIZE is firm size, LOSS is a dummy variable indicating firm-years with reported losses, SKEW is the skewness of earnings, ΔEARN is the change in earnings from the prior year, EVOLAT is earnings volatility, HOR is the forecast horizon in months, FOLLOW is the number of analysts following the firm (per I/B/E/S), and DISP is the forecast dispersion. We also use a series of dummy variables for year (YEAR) and industry classification (INDUSTRY).¹²

Table 3, column 2 presents the results of estimating the forecast accuracy model, Equation (4), using the mean consensus forecasts. We correct the t-values reported in Table 3 for heteroskedasticity using White (1980) standard errors. Standard statistical methods of testing for outliers and influential observations (R-Student, DFBETAS, DFFITS, and Cook's D) show 23 extreme observations. We report the results of the estimations without these observations and after winsorizing the data at the 1/2 percent level. Using the full sample leads to similar inferences, as does truncating at the 1 percent and 3 percent levels.

Our primary result is that in the forecast accuracy model, the estimated international diversification coefficient is negative and significant at the 1 percent level. Thus, greater firm internationalization is associated with less accurate analysts' forecasts, consistent with H1. For the control variables, the coefficient estimates for industrial diversification, firm size, and loss firm-years are insignificantly different from zero. Consistent with Dunn and

¹² We measure industry affiliation using two-digit SIC codes (in parentheses). Our sample spans 62 different industries. The largest industry groups in the sample are: Oil and Gas Extraction (13), Paper and Allied Products (26), Printing and Publishing (27), Chemicals and Allied Products (28), Fabricated Metal Products (34), Industrial Machinery and Equipment (35), Electronic and other Electric Equipment (36), Instruments and Related Products (38), Electric, Gas, and Sanitary Services (49), Wholesale Trade—Durable Goods (50), Depository Institutions (60), and Business Services (73).

TABLE 3

The Association between International Diversification and the Accuracy and Bias (Optimism) in Analysts' Forecasts, after Controlling for Other Determinants of Forecast Accuracy and Bias, for the Period 1995–1998 (n = 3,495)^a

$$\begin{aligned} \text{ACCURACY}_t = & A_0 + A_1(\text{ID}_{t-1}) + A_2(\text{IND}_{t-1}) + A_3(\text{SIZE}_{t-1}) + A_4(\text{LOSS}_t) \\ & + A_5(\text{SKEW}_{t-1}) + A_6(\Delta\text{EARN}_t) + A_7(\text{EVOLAT}_{t-1}) + A_8(\text{HOR}_{t-1}) \\ & + A_9(\text{FOLLOW}_{t-1}) + A_{10}(\text{DISP}_{t-1}) + A_{11}(\text{YEAR}_t) + A_{12}(\text{INDUSTRY}_{t-1}) + e \end{aligned}$$

$$\begin{aligned} \text{BIAS}_t = & B_0 + B_1(\text{ID}_{t-1}) + B_2(\text{IND}_{t-1}) + B_3(\text{SIZE}_{t-1}) + B_4(\text{LOSS}_t) + B_5(\text{SKEW}_{t-1}) \\ & + B_6(\Delta\text{EARN}_t) + B_7(\text{EVOLAT}_{t-1}) + B_8(\text{HOR}_{t-1}) + B_9(\text{FOLLOW}_{t-1}) \\ & + B_{10}(\text{DISP}_{t-1}) + B_{11}(\text{YEAR}_t) + B_{12}(\text{INDUSTRY}_{t-1}) + e' \end{aligned}$$

	Forecast Accuracy		Forecast Bias (Optimism)	
	Predicted	Coefficient	Predicted	Coefficient
	Sign	(t-statistic)	Sign	(t-statistic)
	(1)	(2)	(3)	(4)
Intercept		0.00 (0.06)		0.01 (0.38)
<i>Hypothesized Variable</i>				
Degree of International Diversification (ID)	—	−0.01*** (−5.52)	+	0.01*** (6.41)
<i>Control Variables</i>				
Industry Diversification (IND)	—	−0.00 (−0.53)	+	0.03 (1.06)
Firm Size (SIZE)	?	−0.00 (−0.61)	?	0.00 (0.48)
Loss Firm-Year (LOSS)	—	0.01 (0.81)	+	0.01 (0.07)
Earnings Skewness (SKEW)	—	−0.58*** (−3.80)	+	0.51*** (6.06)
Change in Earnings (ΔEARN)	—	−0.80*** (−10.53)	+	0.86*** (20.49)
Earnings Volatility (EVOLAT)	—	−0.05* (−1.83)	+	0.01 (0.66)
Forecast Horizon (HOR)	—	−0.00** (−1.98)	+	0.00 (0.88)
Number of Analysts Following (FOLLOW)	+	0.01* (1.67)	—	−0.00 (−0.87)
Forecast Dispersion (DISP)	—	−0.09*** (−3.39)	—	−0.08*** (−4.30)
Adjusted R ²		0.85		0.89
F-Value		549.29		815.64

***, **, * Significant at the 1 percent, 5 percent, and 10 percent levels, respectively. The standard errors used to compute the t-values (in parentheses) are corrected for heteroskedasticity using White (1980) standard errors.

^a YEAR is a series of dummy variables that denote the year of the observation. INDUSTRY is a series of dummy variables to denote industry affiliation, based on two-digit SIC codes. We define the remaining variables in Table 2. The YEAR and INDUSTRY dummies are suppressed.

Nathan (2000), we find that firm size and analyst following are highly correlated, and that in our regressions only one of these variables is significant. We also find that industrial diversification is significant in regressions without international diversification (as in Bodnar et al. 1999), but Table 3 shows that industrial diversification becomes insignificant after the inclusion of international diversification. However, earnings skewness, change in earnings, earnings volatility, forecast horizon, and forecast dispersion are all associated with less accurate analysts' forecasts (SKEW, Δ EARN, and DISP at the 1 percent level; HOR at the 5 percent level; and EVOLAT at the 10 percent level). In contrast, the number of analysts following a firm is positively related to consensus forecast accuracy (at the 10 percent level).

Table 3, column 4 reports our results from estimating the bias regression specified in Equation (5), using mean consensus forecasts. Our primary result is that the estimated international diversification coefficient is positive and significant at the 1 percent level, consistent with H2 that greater corporate international diversification is associated with more optimistic analysts' forecasts.¹³ For the control variables, the coefficient estimates for industrial diversification, firm size, and loss firm-years are insignificantly different from zero. We also repeat the tests after excluding loss firms from the analysis, and find that firm internationalization is still positively associated with forecast bias. We also find that both earnings skewness and the change in earnings are associated with more optimistic analysts' forecasts (at the 1 percent level). Furthermore, although the estimates for earnings volatility, forecast horizon, and the number of analysts following a firm are insignificant, the coefficient estimate for dispersion is negative and significant at the 1 percent level.

Our results suggest that firm internationalization is an incremental determinant of forecast accuracy and bias. In the accuracy regressions, all else constant, the firm with the average level of international diversification has a lower forecast accuracy (relative to purely domestic firms) of -0.01 (coefficient estimate multiplied by the average level of international diversification). With a mean forecast accuracy value of -0.05 , this result indicates that the average level of firm international diversification is associated with a 22.1 percent higher-than-average forecast error. We also see that the average level of firm international activity indicates approximately a 22.4 percent higher (more optimistic) forecast bias, relative to that of domestic firms.

V. ADDITIONAL SENSITIVITY TESTS

Earnings Predictability and International Diversification

Das et al. (1998) report that forecast bias is a function of the predictability of earnings, measured using factor analysis of prediction errors based on various time-series models of earnings. To determine whether our results are robust to including the Das et al. (1998) measure of earnings unpredictability in Equations (4) and (5), we estimate the predictability measure described in Das et al. (1998). We measure UNPRED using the prediction errors from five separate time-series models as inputs into the factor analysis. The first three models are based on annual earnings data and consist of a random-walk model, a submartingale model, and an autoregressive model. The last two are based on quarterly data, and include a seasonal random-walk model and a Foster (1977) model. We use the previous ten years of earnings data in estimating the models. We find that international diversification

¹³ Because about 43 percent of the sample is composed of domestic firms, we also conduct the regression tests separately for the firms that report international operations. International diversification is still negatively related to forecast accuracy and positively related to forecast bias (significant at the 1 percent and 5 percent levels, respectively). Furthermore, replacing earnings volatility with the volatility of foreign earnings leads to similar inferences.

remains significantly associated with less accurate and more optimistic analysts' forecasts even after including UNPRED in our accuracy and bias models, Equations (4) and (5).

Alternative Specifications

Although our primary specifications control for forecast horizon, another approach is to use a subset of the sample with the same forecast horizon. Using the subset of firms with December fiscal year-ends, we continue to find that international diversification is associated with less accurate and more optimistic analysts' forecasts. Other subsets with similar forecast horizons yield similar results. Repeating the analysis after creating subsets based on quartiles of firm size and analyst following again leads to similar results, as does repeating the tests year-by-year.

Variance inflation factors and ridge regressions suggest that multicollinearity is unlikely to affect our inferences on firm internationalization. The variance inflation factors in the two primary regressions are close to 1, and within the values where multicollinearity is not a concern (Kennedy 1998, 190). The ridge regression results are similar to the reported results.

VI. SUMMARY AND CONCLUSION

We document that greater corporate international diversification is associated with less accurate and more optimistic analysts' forecasts. This finding holds true even after we control for previously identified determinants of forecast accuracy and bias, such as earnings variability, forecast horizon, firm size, and industrial diversification. Thus, we conclude that firm-level international operations are a significant dimension of forecasting task complexity.

A limitation of this study is that our definition of internationalization implicitly assumes that the effects of expansion outside the United States are the same regardless of the countries into which firms expand. However, the countries in which firms operate might affect the quality of analysts' forecasts. For example, the forecasting task might well be more complex for firms operating in developing countries than for firms operating in countries with more mature market economies.

Another limitation is that we do not control for all possible determinants of accuracy and bias (e.g., selection bias and promoting brokerage-firm business). Also, the SEC's recent Regulation FD could affect the association between international operations and analyst optimism. If analysts biased their forecasts upward to gain favor with managers hoping to obtain nonpublic information during our 1995–1998 sample period, then the new Regulation FD means that managers are now less likely to give private information to analysts regardless of their forecasts, which is likely to attenuate analysts' propensity toward optimism.

REFERENCES

- Abarbanell, J., and R. Lehavy. 2000. Differences in commercial database reported earnings: Implications for inferences concerning analyst forecast rationality, the association between prices and earnings, and firm reporting discretion. Working paper, The University of North Carolina at Chapel Hill.
- Ashbaugh, H., and M. Pincus. 2001. Domestic accounting standards, international accounting standards, and the predictability of earnings. *Journal of Accounting Research* 39 (December): 417–434.
- Balakrishnan, R., T. Harris, and P. Sen. 1990. The predictive ability of geographic segment disclosures. *Journal of Accounting Research* 28 (Autumn): 305–325.

- Baldwin, B. 1984. Segment earnings disclosure and ability of security analysts to forecast earnings per share. *The Accounting Review* 59 (July): 376–389.
- Bamber, L., O. Barron, and T. Stober. 1997. Trading volume and different aspects of disagreement coincident with earnings announcements. *The Accounting Review* 72 (October): 575–597.
- Basu, S., L. Hwang, and C. Jan. 1998. International variation in accounting measurement rules and analysts' earnings forecast errors. *Journal of Business Finance & Accounting* 25 (November/December): 1207–1248.
- Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics* 11 (July): 255–274.
- Bodnar, G., and J. Weintrop. 1997. The valuation of the foreign income of U.S. multinational firms: A growth opportunities perspective. *Journal of Accounting and Economics* 24 (December): 69–97.
- , C. Tang, and J. Weintrop. 1999. Both sides of corporate diversification: The value impacts of geographical and industrial diversification. Working paper, University of Pennsylvania, Philadelphia, PA.
- Brown, L. D., R. Hagerman, P. Griffin, and M. Zmijewski. 1987. Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *Journal of Accounting and Economics* 9 (April): 61–87.
- , G. Richardson, and S. Schwager. 1987. An information interpretation of financial analyst superiority in forecasting earnings. *Journal of Accounting Research* 25 (Spring): 49–67.
- . 1993. Earnings forecasting research: Its implications for capital market research. *International Journal of Forecasting* 9 (November): 295–320.
- . 1997. Analyst forecasting errors: Additional evidence. *Financial Analyst Journal* 53 (November/December): 81–88.
- . 2001. A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research* 39 (September): 221–242.
- Brown, P., G. Foster, and E. Noreen. 1985. *Security Analyst Multi-Year Earnings Forecasts and the Capital Market*. Sarasota, FL: American Accounting Association.
- Clement, M. 1999. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics* 21 (June): 285–303.
- Collins, D. 1976. Predicting earnings with subentity data: Some further evidence. *Journal of Accounting Research* 14 (Spring): 163–177.
- Das, S., C. Levine, and K. Sivaramakrishnan. 1998. Earnings predictability and bias in analysts' earnings forecasts. *The Accounting Review* 73 (April): 277–294.
- , and S. Saudagaran. 1998. Accuracy, bias, and dispersion in analysts' earnings forecasts: The case of cross-listed foreign firms. *Journal of International Financial Management and Accounting* 9 (February): 16–33.
- Dechow, P., A. Hutton, and R. Sloan. 2000. The relation between analysts' long-term earnings forecasts and stock price performance following equity offerings. *Contemporary Accounting Research* 17 (Spring): 1–32.
- DeGeorge, F., J. Patel, and R. Zeckhauser. 1999. Earnings manipulation to exceed thresholds. *Journal of Business* 72 (January): 1–33.
- Downen, R. J. 1996. Analyst reaction to negative earnings for large well-known firms. *Journal of Portfolio Management* 21 (Fall): 49–55.
- Dugar, A., and S. Nathan. 1995. The effect of investment banking relationships on financial analysts' earnings forecasts and investment recommendations. *Contemporary Accounting Research* 12 (Fall): 131–160.
- Dunn, K., and S. Nathan. 2000. The effect of industry diversification on analysts' earnings forecasts. Working paper, Baruch College–CUNY, New York, NY.
- Easterwood, J. C., and S. R. Nutt. 1999. Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *Journal of Finance* 54 (October): 1777–1797.
- Financial Accounting Standards Board (FASB). 1976. *Financial Reporting for Segments of a Business Enterprise*. Statement of Financial Accounting Standards No. 14. Stamford, CT: FASB.

- . 1997. *Disclosures about Segments of an Enterprise and Related Information*. Statement of Financial Accounting Standards No. 131. Norwalk, CT: FASB.
- Foster, G. 1977. Quarterly accounting data: Time-series properties and predictive-ability results. *The Accounting Review* 52 (January): 1–21.
- Gay, G., and J. Nam. 1998. The underinvestment problem and corporate derivative use. *Financial Management* 27 (Spring): 53–69.
- Goldberg, S., and F. Hefflin. 1995. The association between the level of international diversification and risk. *Journal of International Financial Management and Accounting* 6 (Spring): 1–25.
- Gu, Z., and J. Wu. 2001. Earnings skewness and analysts' forecast bias. Working paper, Carnegie Mellon University, Pittsburgh, PA.
- Haw, I., K. Jung, and W. Ruland. 1994. The accuracy of financial analysts forecasts after mergers. *Journal of Accounting, Auditing, and Finance* 9 (Summer): 465–486.
- Hayes, R. M. 1998. The impact of trading commission incentives on stock coverage and earnings forecast decisions by security analysts. *Journal of Accounting Research* 36 (Autumn): 299–320.
- Herrmann, D., and W. Thomas. 2000. An analysis of segment disclosures under SFAS No. 131 and SFAS No. 14. *Accounting Horizons* 14 (September): 287–302.
- Hwang, L., C. Jan, and S. Basu. 1996. Loss firms and analysts' earnings forecast errors. *Journal of Financial Statement Analysis* 1 (Winter): 18–30.
- Kang, S., J. O'Brien, and K. Sivaramakrishnan. 1994. Analysts' interim earnings forecasts: Evidence on the forecasting process. *Journal of Accounting Research* 32 (Spring): 103–112.
- Kennedy, P. 1998. *A Guide to Econometrics*. 5th edition. Cambridge, MA: The MIT Press.
- Kinney, W. 1971. Predicting earnings: Entity vs. sub-entity data. *Journal of Accounting Research* 9 (Spring): 127–142.
- Kogut, B. 1983. Foreign direct investment as a sequential process. In *The Multinational Corporations in the 1980s*, edited by C. Kindleberger. Cambridge, MA: MIT Press.
- Kross, W., B. T. Ro, and D. Schroeder. 1990. Earnings expectations: The analysts' information advantage. *The Accounting Review* 65 (April): 461–476.
- Lang, M., and R. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 (October): 467–492.
- Lim, T. 2001. Rationality and analysts' bias. *Journal of Finance* 56 (February): 369–385.
- Lin, H., and M. McNichols. 1998. Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics* 20 (February): 101–127.
- Lys, T., and L. Soo. 1995. Analysts' forecast precision as a response to competition. *Journal of Accounting, Auditing, and Finance* 10 (Fall): 751–763.
- Matsumoto, D. A. 2002. Management's incentive to avoid negative earnings surprises. *The Accounting Review* 77 (July) (forthcoming).
- McNichols, M., and P. C. O'Brien. 1997. Self selection and analyst coverage. *Journal of Accounting Research* 35 (Autumn): 167–199.
- Nichols, D., L. Tunnel, and C. Seipel. 1995. Earnings forecast accuracy and geographic segment disclosures. *Journal of International Accounting, Auditing and Taxation* 8: 113–126.
- O'Brien, P. C., and R. Bhushan. 1990. Analyst following and institutional ownership. *Journal of Accounting Research* 28 (Spring): 55–82.
- Reeb, D., C. Kwok, and Y. Baek. 1998. Systematic risk in the multinational corporation. *Journal of International Business Studies* 29 (Second Quarter): 263–279.
- Richardson, S., S. Teoh, and P. Wysocki. 2001. The walkdown to beatable analyst forecasts: The role of equity issuance and insider trading incentives. Working paper, University of Michigan, Ann Arbor, MI.
- Roberts, C. 1989. Forecasting earnings using geographic segment data: Some U.K. evidence. *Journal of International Financial Management and Accounting* 1 (Summer): 130–151.
- Schipper, K. 1991. Analysts' forecasts. *Accounting Horizons* 5 (December): 105–119.
- Shapiro, A. 1978. Financial structure and the cost of capital in the multinational corporation. *Journal of Financial and Quantitative Analysis* 13 (June): 211–266.

- Sullivan, D. 1994. Measuring the degree of internationalization of a firm. *Journal of International Business Studies* 25 (Second Quarter): 325–334.
- Thomas, W. 2000. A test of the market's mispricing of domestic and foreign earnings. *Journal of Accounting and Economics* 28 (December): 243–267.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (May): 817–838.
- Wiedman, C. 1996. The relevance of characteristics of information environment in the selection of a proxy for the market's expectations for earnings: An extension of Brown, Richardson, and Schwager [1987]. *Journal of Accounting Research* 34 (Autumn): 313–324.
- World Trade Organization. 2000. *WTO Annual Report*. Geneva, Switzerland: WTO Publications.

Economic Determinants of Audit Committee Independence

April Klein
New York University

ABSTRACT: This paper provides empirical evidence that audit committee independence is associated with economic factors. I find that audit committee independence increases with board size and board independence and decreases with the firm's growth opportunities and for firms that report consecutive losses. In contrast, no relation is found between audit committee independence and creditors' demand for accounting information. Although the analyses are based on data from 1991 to 1993, these results have implications for NYSE and NASDAQ listing requirements for audit committees adopted in December 1999. Specifically, the new requirements give firms the option of including non-outside directors on their audit committees if it is in the best interests of the firm to do so.

Keywords: *audit committee; outside directors; corporate governance; board of directors.*

Data Availability: *Data used for this study are derived from a proprietary database.*

Just as "one size doesn't fit all" when it comes to board governance, "one size can't fit all" audit committees. Within broad parameters, each audit committee should evolve and develop its own guidelines suited to itself and its corporation.

—New York Stock Exchange and National Association of Securities Dealers (1999)

I. INTRODUCTION

In December 1999, in response to the SEC's call for improving the effectiveness of corporate audit committees in overseeing the financial-reporting process (Levitt 1998), the NYSE and NASDAQ modified their listing requirements for large U.S. companies. Under the new standards, firms must maintain audit committees with at least three directors, "all of whom have no relationship to the company that may interfere with the exercise of their independence from management and the company" (*NYSE Listed Company Manual*

Financial support is provided by a summer grant from the Stern School of Business. I thank Ashiq Ali, Eli Bartov, Stephen Bryan, Eric Koh, Baruch Lev, James Ohlson, Jeffrey Simonoff, the editor (Michael Bamber), two anonymous referees, and the NYU Accounting and Statistics Departments for their constructive comments.

Submitted April 2000
Accepted November 2001

§303.01[B][2][a]). Although this statement implies that firms must maintain audit committees with outside members only, listing requirements provide for the appointment of certain affiliated directors if the board determines it is in the best interests of the corporation for these individuals to serve on its audit committee (see NYSE Rule §303.01[B][3][b] and NASDAQ Rule 4310[c][26][B][ii]). Thus, firms have some flexibility in determining audit committee composition.

Because enforcement standards are constantly evolving, one question of interest is to what degree the SEC and the exchanges should allow firms to exercise this option. I provide insights into this question by examining economic determinants behind differences in audit committee independence for a sample of more than 400 large U.S. firms that were publicly traded during most of 1991–1993. During this time, exchange rules allowed more flexibility with respect to audit committee independence and many firms opted for audit committees with less than 100 percent outside directors (Vicknair et al. 1993; Verschoor 1993; Klein 1998, 2001; Parker 2000).

I develop predictions and test determinants of audit committee independence. I expect and find that the percentage of outside directors on the audit committee is limited by board size and overall board independence. Firms incur costs in expanding the board to include more outside directors (e.g., Yermack 1996) and in enlisting outside directors instead of inside directors who have firm-specific knowledge (e.g., Fama and Jensen 1983; Klein 1998). Thus, audit committee independence is costly to the firm.

I expect the demands for audit committee independence to emanate from management, shareholders, and creditors. Consistent with my expectation that managers' demand for directors with inside expertise increases in proportion to the complexities and uncertainties associated with growth opportunities (Williamson 1975), I find that audit committee independence decreases with the firm's growth opportunities. Consistent with the hypothesis that shareholders' demand for accurate, unbiased financial accounting data depends on the published financial accounting data's potential informativeness for equity valuation, I find that firms that have experienced two or more consecutive losses (which typically are less value-relevant than positive profits) have less independent audit committees. However, I find no evidence that audit committee independence is associated with the degree of debt in the firm's capital structure. This result is inconsistent with the expectation that creditors' demand for unbiased accounting data for use in debt covenants increases with debt.¹

One implication of these findings is that firms tailor audit committee composition to suit their specific economic environments. This suggests that the SEC and the stock exchanges may wish to continue to allow firms some flexibility to include non-outside directors on their audit committees.

II. THE AUDIT COMMITTEE

The Audit Committee's Role as Monitor of the Firm's Financial-Reporting Process

By state law, boards of directors may conduct their work through the full board or delegate their authority to standing committees responsible to the board—for example, Delaware General Corporate Law §141(c) allows boards to set up committees. One such committee is the audit committee.

The audit committee provides, on behalf of the board of directors, oversight responsibility for the firm's financial-reporting process. According to the *Blue Ribbon Committee*

¹ Creditors also use accounting numbers to assess firm value in liquidation (as in the abandonment option). See, for example, Barth et al. (1998).

Report, the audit committee is the “the ultimate monitor” of the financial accounting reporting system (NYSE and NASD 1999, 7). The audit committee selects the outside auditor and meets separately with senior financial management and with the external auditor. The committee also questions management, internal auditors, and external auditors to determine whether they are acting in the firm’s best interests.²

Consistent with the *Blue Ribbon Committee Report* and with prior studies (e.g., Carcello and Neal 2000), I assume that audit committee members who are independent of management are better monitors of the firm’s financial accounting process.³ Benefits of effective monitoring include transparent financial statements, active trading markets, and the ability to use unbiased financial accounting numbers as inputs into contracts among shareholders, senior claimants, and management.

NYSE and NASDAQ Audit Committee Requirements

Before December 1999, stock exchanges’ and NASDAQ’s rules for audit committee composition were vague at best. They required or encouraged large, U.S.-listed companies to maintain audit committees with all or a majority of their members independent of management. However, the listing requirements did not define “independence.”⁴

In December 1999, the NYSE and NASDAQ modified their audit committee requirements, based on their *Report and Recommendations of the Blue Ribbon Committee on Improving the Effectiveness of Corporate Audit Committees*, which they had issued in the previous February. Simultaneously, the SEC adopted new rules to improve disclosures related to audit committees (see SEC Release No. 34-42266, *Adopting Rules Regarding Disclosure by Audit Committees, Including Discussions with Auditors Regarding Financial Statements*). All large, U.S.-listed companies must now maintain audit committees with at least three independent directors.⁵ Both exchanges disallow directors from serving on the audit committee if they are current employees, have been employees within the last three years, have cross-compensation committee links, or are immediate family members of an executive officer. In addition, the NASDAQ bars from the audit committee directors who accept more than \$60,000 in non-director compensation from the firm, or whose employers received at least \$200,000 from the firm in any of the three past years.

However, both the NYSE and NASDAQ allow the firm to appoint to the audit committee directors who have business relationships with the firm if the board determines that the individual’s membership on the committee is in the corporation’s best interests. NASDAQ Rule 4310(c)(26)(B)(ii) allows the board under “limited circumstances” to appoint one former employee or family member to the audit committee if the board determines this is in the best interest of the corporation and its shareholders. NYSE §303.01(B)(3)(b) gives the board greater discretion. If the board determines that the business relationship does not

² For example, see *Audit Committee Effectiveness—What Works Best*, a guide for audit committees written by PricewaterhouseCoopers (2000), under the aegis of the Institute of Internal Auditors Research Foundation.

³ Fama and Jensen (1983) and Fama (1980) also argue that outside directors have incentives to effectively monitor top management.

⁴ Advocacy of independent audit committees has a long tradition. In 1940, the SEC first recommended that firms establish audit committees with only nonofficer board members (Accounting Series Release No. 19). In October 1987, the Treadway Commission advocated that audit committees include only independent directors (National Commission on Fraudulent Financial Reporting 1987).

⁵ See *NYSE Listed Company Manual*, New York Stock Exchange, §303.01(B)(2)(a), and *The NASDAQ Stock Market Listing Requirements*, National Association of Securities Dealers, §4310(c)(26)(B). See also SEC Release Nos. 34-42231, 34-42232, and 34-42233, *Adopting Changes to Listing Requirements for the NASD, AMEX, and NYSE Regarding Audit Committees*.

compromise the individual's independence, then that director may serve on the board's audit committee. Thus, despite recent changes in exchange requirements, audit committee independence is still an open issue.

III. DETERMINANTS OF AUDIT COMMITTEE INDEPENDENCE

Board Size and Board Independence

I develop hypotheses relating audit committee independence to the supply of available outside directors. Because the audit committee is a subset of the full board, its composition depends on the board's overall structure.

Lipton and Lorsch (1992), Jensen (1993), and Yermack (1996) argue that the board's decision-making quality decreases with board size because the more people in the group, the lower the group's coordination and processing skills (see Steiner 1972; Hackman 1990). Yermack (1996) finds evidence consistent with this argument. Specifically, Tobin's Qs, several accounting profitability ratios, and CEO turnover rates are negatively related to board size.

If the firm limits board size, then the number of directors available to serve on the audit committee also will be limited. The typical board has between three and six board committees to staff, with each committee having at least two members (Klein 1998). Two committees, the executive compensation committee and the nominating committee, require independent directors to ensure monitoring. Thus, a smaller board may have to choose one or more nonindependent directors to serve on the audit committee to meet overall committee staffing needs.

Stated in the alternative form:

H1: Audit committee independence is positively related to the number of directors sitting on the entire board.

The larger the pool of outside directors on the board, the easier it is for the board to have an independent audit committee. However, boards require both outside and non-outside directors to fulfill their duties. Outside directors serve as monitors and help alleviate agency conflicts between shareholders and upper management. Inside and affiliated directors have the specialized expertise about the firm's activities to evaluate and ratify its future strategic plans (Williamson 1975; Fama and Jensen 1983). Consistent with this argument, Klein (1998) finds that the percentage of inside directors on board investment or finance committees is positively associated with firm value. Thus, board independence reflects the trade-off between director independence and director expertise, which, in turn, reflects the balancing of the firm's monitoring needs and its requirements for specialized information.

If board independence varies across firms, then, in the alternative form:

H2: Audit committee independence is positively related to board independence.

The Demands for Audit Committee Independence

Growth Opportunities

Williamson (1975) and Fama and Jensen (1983) argue that a firm structures its board in response to its needs to obtain unbiased, expert information. Both papers assert that a board's demand for knowledgeable directors increases with the firms' complexity and uncertainties. Inside directors have direct knowledge about their firm's operations and investment horizons. Affiliated directors or their companies have ties with the firm and often provide expertise to the firm about suppliers, customers, financial opportunities, or legal

issues. Because of the complexities and uncertainties associated with growth opportunities, I expect high-growth opportunity firms' managers and shareholders to demand less independent boards, resulting in less independent audit committees.

In the alternative form:

H3: Audit committee independence is negatively related to the firm's expected growth in earnings or cash flows.

Consecutive Losses

Hayn (1995), Lipe et al. (1998), Amir et al. (1999), and Collins et al. (1999) show that the cross-sectional returns (or price) earnings relation is much weaker for firms reporting losses than for firms reporting profits. In addition, Hayn (1995) reports negative coefficients on the regression of returns on earnings, with R^2 values near 0.0 percent, for her sample of firms posting losses over two or more consecutive years. These studies suggest that financial statements on the whole are less value-relevant for firms suffering repeated losses than for profitable firms. Thus, I expect shareholders of firms with past consecutive losses to demand less scrutiny of the financial-reporting system and, consequently, to have a lower demand for audit committee independence.

In the alternative form:

H4: Audit committee independence is lower for firms reporting a series of consecutive, past losses.

Creditors

Creditors write debt contracts that contain accounting-based covenants to monitor management and shareholders (Jensen and Meckling 1976; Smith and Warner 1979; Leftwich 1983; Watts and Zimmerman 1990). However, managers sometimes manipulate earnings to delay or avoid debt covenant violations. For example, DeFond and Jiambalvo (1994) conclude that managers overstate earnings in the year before debt covenant violations. Thus, creditors' demand for audit committee independence should increase with the firm's debt-to-assets ratio due to their increased demands for monitoring the integrity of the firms' financial accounting reports.

In the alternative form:

H5: Audit committee independence is positively related to the firm's debt-to-assets ratio.

CEO on Compensation Committee

Although exchange and NASDAQ regulations restrict the CEO from serving on the board's audit committee, no regulations prohibit the CEO from sitting on the board's executive compensation committee. I make no prediction, *ex ante*, about the sign of the relation between the CEO serving on the compensation committee and audit committee independence. On the one hand, a board that allows its CEO to sit on its executive compensation committee may be CEO-friendly, thus providing less monitoring of the firm's financial accounting reporting process. For example, Klein (2001) finds that earnings management is higher in firms where the CEO serves on the board's compensation committee. On the other hand, shareholders may want to curb the CEO's ability to distort financial statements by demanding greater audit committee independence. Thus, Hypothesis 6 is nondirectional:

H6: Audit committee independence is different if the CEO sits on the board's executive compensation committee than if the CEO does not.

Substitute Monitoring Mechanisms

Nonmanagement blockholders (or their representatives) sometime serve on the audit committee, giving them the opportunity to monitor the firm's financial-reporting process. Klein (1998, 2001) finds that firms with large, non-inside stockholders on audit committees are more productive and exhibit less earnings management. If large, nonmanagement blockholders on the audit committee actively monitor the financial-reporting process, then there is less need for outside directors on the audit committee.

In the alternative form:

H7: Audit committee independence is lower if a large, non-inside shareholder sits on the board's audit committee.

Jensen and Meckling (1976) argue that directors' shareholdings act as a monitoring device. If outside director shareholdings substitute for outsiders on the audit committee, then in the alternative form:

H8: Audit committee independence is negatively related to the percentage of shares held by outside directors.

Firm Size

I also control for firm size. Larger firms have stronger internal controls systems than smaller firms (O'Reilly et al. 1998). If the firms' internal controls act as in-house monitoring mechanisms, then larger firms require less alternative monitoring of their reporting systems and therefore need lower levels of audit committee independence. Alternatively, if shareholders are more apt to sue larger firms for misstated or fraudulent financial statements, then larger firms may try to inoculate themselves against lawsuits by adopting stronger monitoring mechanisms, such as greater audit committee independence. Thus, the association between firm size and audit committee independence is indeterminate.

IV. RESEARCH DESIGN

Sample Selection

I hand-collected data about boards and board audit committees from SEC-filed proxy statements. Initially, I include all U.S. firms listed on the S&P 500 as of March 31, 1992 and 1993, with annual shareholder meetings between July 1, 1991, and June 30, 1993. I exclude 65 bank and financial institution firm-years and 38 insurance company firm-years. I also delete three firm-years with missing audit committee composition data, and 58 firm-years with missing Compustat or CRSP data. The final sample has 803 firm-year observations.

Schedule 14A (the proxy statement) requires firms to disclose each director's name, current directorships; family relationships between each and any director, nominee or executive officer; significant current or proposed transactions with management; significant business relationships with the firm, current firm shareholdings, and business experience during the last five years.⁶ Schedule 14A (Item 7[e][1]) requires firms to state whether they

⁶ Items 404(a) and 404(b) of Regulation S-K of the 1934 Securities and Exchange Act define significant business transactions. Item 404(a) specifies a threshold of \$60,000 for a transaction to be considered significant. Item 404(b) defines "certain business relationships" to include significant payments to the firm in return for services or property, significant indebtedness by the firm, outside legal counseling, investment banking, consulting fees, and other joint ventures.

have standing audit, compensation, or nominating board committees. If such committees exist, then firms must disclose their functions and responsibilities, their members, and the number of times each committee met during the last fiscal year.

I define *outside directors* as having no affiliation with the firm other than serving as directors. Consistent with NYSE and NASD (1999) and with exchange definitions, *affiliated directors* include former employees, relatives of the CEO, those with significant transactions or business relationships with the firm as defined by Items 404(a) and (b) of Regulation S-K, and those on interlocking boards. *Inside directors* are current employees.

Regression Model

I measure the associations between audit committee independence and the explanatory variables by estimating the following regression:

$$\begin{aligned} \%Audout = & \alpha + \beta_1 \text{Board Size} + \beta_2 \%Outsiders + \beta_3 \text{Growth Opportunities} \\ & + \beta_4 \text{Losses} + \beta_5 \text{Debt-to-Assets} + \beta_6 \text{CEO on Compensation Committee} \\ & + \beta_7 5\% \text{Blockholder on Audit Committee} + \beta_8 \% \text{Outside Director Holdings} \\ & + \beta_9 \text{Firm Size} + \mu. \end{aligned}$$

$\%Audout$ is the logistical transformation of the percentage of outsiders on the audit committee.⁷ Board Size is the natural log of the number of board members (Yermack 1996) and $\%Outsiders$ is the logistical transformation of the percent of outside directors on the board.⁸ I obtain data for these three variables from the shareholders' meeting proxy statement. Following Smith and Watts (1992), I define Growth Opportunities as the three-year market value of equity plus the book value of liabilities divided by the three-year book value of assets, all ending on the fiscal year prior to the shareholders' meeting.⁹ This metric captures the proportion of firm value represented by growth opportunities relative to assets in place. Losses equals 1 if the firm reported losses for each of the two years prior to the firm's shareholders' meeting (Compustat item 18 is negative for fiscal years t and $t-1$), and 0 otherwise. Losses captures the diminished value-relevance of financial reporting for firms experiencing extended periods of losses. Debt-to-Assets is the three-year long-term debt-to-assets ratio at the end of the fiscal year prior to the shareholders' meeting (three-year Compustat item 9 divided by three-year item 6). CEO on Compensation Committee is a dichotomous variable set to 1 if the CEO sits on the board's compensation committee, and 0 otherwise. 5%Blockholder on Audit Committee is a dichotomous variable set to 1 if a nonmanagement director holding at least 5 percent of the firm's shares sits on the board's audit committee, and 0 otherwise.¹⁰ %Outside Director Holdings is the percentage of shares held by outside directors. The latter two items are in the firm's proxy statement. Firm Size is the natural log of the firm's assets (Compustat item 6) at the end of the fiscal year prior to the shareholders' meeting.

⁷ The logistical transformation is $\ln(\%Audout/(1-\%Audout)+1)$. I use this transformation because the values of $\%Audout$ are confined to the interval from 0 to 1, whereas the logistically transformed values extend from $-\infty$ to $+\infty$. Thus, an intrinsically non-normal distribution is transformed into a more normal distribution.

⁸ I use the latter transformation to be internally consistent with the transformed $\%Audout$ variable. Results based on untransformed values as well as the natural log of $\%Outsiders$ are qualitatively the same as those reported in the text and are not shown separately.

⁹ The numerator is Compustat items $(6 - 60 + (24 \times 25))$ summed over fiscal years t , $t-1$, and $t-2$. The denominator is Compustat item 6 summed over fiscal years t , $t-1$, and $t-2$.

¹⁰ I would like to thank Lee-Seok Hwang for these data.

V. EMPIRICAL RESULTS

Descriptive Statistics

Table 1 reports data on board and audit committee composition. Consistent with other board composition studies or surveys, 58.4 percent of the board is outsiders, less than one quarter (22.5 percent) is insiders, and the rest (19.1 percent) are affiliated directors.¹¹ In contrast, the audit committee includes a preponderance of outsiders (79.6 percent) and few insiders (1.4 percent), but about the same percentage of affiliated directors (19.0 percent). Because of these affiliated directors, 43.4 percent of firms have audit committees with outside directors only, and 86.7 percent have a majority of independent directors.

Table 2 presents the mean, median, and 1st and 3rd quartiles of the untransformed variables used in the regression analysis. The interquartile range for board size is 10–14 members. The interquartile range for %Outsiders is 50.0 percent to 70.6 percent, consistent with Fama and Jensen's (1983) argument that boards should include some insiders and affiliated directors for their expertise.

Growth Opportunities and the Debt-to-Assets ratio have means of 1.40 and 0.27, respectively. 4.6 percent of firms report losses in two or more consecutive years; 9.1 percent of firms allow the CEO to sit on the executive compensation committee; and 5.7 percent

¹¹ Yermack (1996) reports 54 percent outsiders for his sample of 452 firms listed on the *Forbes* 500 between 1934 and 1991. Bhagat and Black (1999) report similar percentages of insiders for 957 large U.S. firms in 1991.

TABLE 1
Composition of Overall Boards of Directors and Audit Committees^a

	(1) <i>Overall Boards of Directors</i>	(2) <i>Audit Committee</i>
<i>Percentage of Directors Who Are:</i>		
Insiders	22.5	1.4
Outsiders	58.4	79.6
Affiliates	19.1	19.0
<i>Percentage of Firms that Have at Least One Member Who Is:</i>		
Inside Director	99.7	4.8
CEO	99.0	1.9
Outside Director	99.0	97.9
Affiliated Director	85.5	54.3
Relative of CEO	11.3	4.3
Former Employee of Firm	51.5	12.3
<i>Percentage of Firms that Have at Least 51% Outside Directors</i>	73.8	86.7
<i>Percentage of Firms that Have 100% Outside Directors</i>	0	43.4

^a Sample is for 803 U.S. firm-years with audit committees listed on the S&P 500 as of March 31, 1992 and 1993 with annual shareholder meetings between July 1, 1991 and June 30, 1993. Banks, financial institutions, insurance companies, and firms with missing Compustat or CRSP data are excluded.

TABLE 2
Descriptive Statistics for Untransformed Variables Used in the Analysis of Audit Committee Independence for a Pooled Sample of 803 Firm-Years 1991–1993

<i>Variable^a</i>	<i>Hypothesis Number^b</i>	<i>Mean</i>	<i>Median</i>	<i>1st Quartile</i>	<i>3rd^d Quartile</i>
%Audout		79.7%	80.0%	66.7%	99.0%
Board Size	1	12.0	12.0	10.0	14.0
%Outsiders	2	58.4%	60.0%	50.0%	70.6%
Growth Opportunities	3	1.40	1.29	1.11	1.61
%Losses	4	4.6%	0	0	0
Debt-to-Assets	5	0.27	0.27	0.16	0.36
CEO on Compensation Committee	6	9.1%	0	0	0
5% Blockholder on Audit Committee	7	5.7%	0	0	0
%Outside Director Holdings	8	1.58%	0.47%	0.15%	1.82%
Assets (\$ million)	Control Variable	8,526	3,190	1,330	8,554

^a Variable definitions:

- %Audout = the percent of outside directors on the audit committee;
- Board Size = the number of board members;
- %Outsiders = the percent of outside directors on the board;
- Growth Opportunities = the three-year market value of the total firm (Compustat items (6 – 60 + (24 × 25)) divided by three-year assets-in-place (Compustat item 6) ending on the fiscal year prior to the shareholders' meeting;
- %Losses = the percent of firms that reported losses (Compustat item 18) in each of the two years prior to the shareholders' meeting;
- Debt-to-Assets = the three-year ratio of the book value of debt (Compustat item 9) divided by the book value of assets (Compustat item 6) ending on the fiscal year prior to the shareholders' meeting;
- CEO on Compensation Committee = 1 if the CEO sits on the board's compensation committee, and 0 otherwise;
- 5% Blockholder on Audit Committee = 1 if a non-inside director with at least 5 percent of the firm's shares sits on the audit committee, and 0 otherwise;
- %Outside Director Holdings = the percentage of shares owned by all outside directors; and
- Assets = the firm's book value of assets (Compustat item 6).

^b Hypothesis number refers to the hypotheses in Section III.

include a non-inside director/blockholder on the audit committee. Outside director shareholdings have an interquartile range of 0.15 percent to 1.82 percent, with a mean of 1.58 percent.

Correlations

Table 3 presents Pearson correlations between the transformed dependent and independent variables. Spearman correlations yield similar results. The following correlations support several hypotheses: %Audout is significantly positively correlated with Board Size

TABLE 3
Pearson Correlations (and p-values) among the Dependent and Explanatory Variables^a

	<u>%Audout</u>	<u>Board Size</u>	<u>%Outsiders</u>	<u>Growth Opportunities</u>	<u>Losses</u>	<u>Debt-to-Assets</u>	<u>CEO on Compensation Committee</u>	<u>5% Blockholder on Audit Committee</u>	<u>%Outside Director Holdings</u>
Board Size	0.06 (0.10)								
%Outsiders	0.50 (0.01)	0.09 (0.01)							
Growth Opportunities	-0.12 (0.01)	-0.16 (0.01)	-0.21 (0.01)						
Losses	-0.03 (0.44)	-0.01 (0.81)	0.02 (0.51)	-0.11 (0.01)					
Debt-to-Assets	0.04 (0.24)	0.19 (0.01)	0.13 (0.01)	-0.30 (0.01)	0.08 (0.03)				
CEO on Compensation Committee	-0.17 (0.01)	-0.15 (0.01)	-0.25 (0.01)	0.20 (0.01)	-0.03 (0.42)	-0.15 (0.01)			
5% Blockholder on Audit Committee	-0.16 (0.01)	-0.01 (0.84)	-0.06 (0.08)	0.07 (0.06)	-0.03 (0.42)	-0.09 (0.02)	0.03 (0.34)		
%Outside Director Holdings	0.05 (0.20)	-0.01 (0.69)	0.09 (0.01)	0.06 (0.08)	0.03 (0.33)	0.02 (0.56)	-0.02 (0.55)	0.36 (0.01)	
Firm Size	0.03 (0.47)	0.54 (0.01)	0.12 (0.01)	-0.25 (0.01)	-0.05 (0.14)	0.31 (0.01)	-0.17 (0.01)	-0.08 (0.03)	-0.09 (0.02)

^a Variable definitions:

%Audout = $\ln(\%Audout / (1 - \%Audout) + 1)$;

Board Size = the natural log of the number of board members; and

%Outsiders = $\ln(\%Outsiders / (1 - \%Outsiders) + 1)$.

See Table 2 for other variable definitions.

(H1) and %Outsiders (H2). %Audout is significantly negatively correlated with Growth Opportunities (H3), CEO on Compensation Committee (H6), and 5%Blockholder on Audit Committee (H7). In contrast, the correlations between audit committee independence and Losses (H4), Debt-to-Assets (H5), %Outside Director Holdings (H8), and Firm Size are insignificantly different from zero.

Table 3 also reveals that many explanatory variables are significantly correlated with each other. The formal hypothesis tests are based on multiple regression analysis.

Multiple Regression Results

The data encompass two consecutive years, with many firms represented in both years. OLS error terms will suffer from serial correlation, resulting in unbiased coefficients but understated standard errors. I use Froot's (1989) procedure to adjust the variance-covariance matrix for dependence among observations from the same firm.

Table 4 presents the empirical results. %Audout is significantly positively associated with Board Size ($p < 0.01$) and with %Outsiders ($p < 0.01$). These findings are consistent with H1 and H2, supporting the view that audit committee independence increases with the supply of available, outside directors.

%Audout is significantly negatively associated with Growth Opportunities ($p < 0.10$), Losses ($p < 0.05$), 5%Blockholder on Audit Committee ($p < 0.01$), and Firm Size ($p < 0.05$). These findings are consistent with H3, H4, and H7, with Firm Size serving as a control variable. The negative coefficient on Growth Opportunities supports the view that managers of growth firms require more board members with specific expertise about their firms. The negative coefficient on Losses is consistent with shareholders demanding lower audit committee independence when financial statements are less value relevant. The negative coefficients on 5%Blockholder on Audit Committee and on Firm Size suggest that large nonmanagement blockholders and strong internal controls substitute for audit committee independence. The coefficients on the debt-to-assets ratio (H5), CEO on Compensation Committee (H6), and Outside Director Holdings (H8) are insignificantly different from zero.

In summary, audit committee independence increases with the supply of available, independent directors and with the demand for monitoring, and decreases with the availability of substitute monitoring mechanisms. %Outsiders plays the biggest role in explaining audit committee independence. A simple regression of %Audout on %Outsiders yields an R^2 of 0.22, compared with the R^2 of 0.25 from the full multiple regression. Nonetheless, the other factors collectively play a significant incremental role in explaining the percentage of outsiders on the audit committee variation in audit committee independence; testing for the differential in R^2 s yields an F-statistic of 5.07, significant at the 0.01 level.

Additional Tests

Are Growth Opportunities Primarily Capturing Hi-Tech Firms?

The variable, Growth Opportunities, may capture disproportionate numbers of firms in high-tech industries or with high research and development expenditures.¹² Reingold (1999) reports that according to Korn/Ferry and Spencer Stuart, high-tech companies such as Internet firms have fewer outsiders on their boards and smaller board sizes than do other types of companies.

¹² Contrary to this conjecture, Smith and Watts (1992) test and find that R&D expenditures cannot substitute for this measure of growth opportunities in their analyses. Lev and Zarowin (1999) also find no relation between R&D expenditures and future growth.

TABLE 4
Explanators of Audit Committee Independence Based on Time-Series Adjusted Regressions^a
for a Pooled Sample of 803 Firm-Years 1991–1993

$$\begin{aligned} \%Audout = & \alpha + \beta_1 \text{Board Size} + \beta_2 \%Outsiders + \beta_3 \text{Growth Opportunities} \\ & + \beta_4 \text{Losses} + \beta_5 \text{Debt-to-Assets} + \beta_6 \text{CEO on Compensation Committee} \\ & + \beta_7 \% \text{Blockholder on Audit Committee} + \beta_8 \% \text{Outside Director Holdings} \\ & + \beta_9 \% \text{Firm Size}^b \end{aligned}$$

<i>Independent Variables</i>	<i>Predicted Sign</i>	<i>Hypothesis Number</i>	<i>Coefficients and (t-statistics)</i>
Intercept			0.39 (0.56)
Board Size	+	1	0.78 (2.62)***
%Outsiders	+	2	1.89 (13.50)***
Growth Opportunities	–	3	–0.06 (–1.86)*
Losses	–	4	–0.55 (–2.32)**
Debt-to-Assets	+	5	–0.26 (–0.59)
CEO on Compensation Committee	+ / –	6	–0.25 (–1.39)
5% Blockholder on Audit Committee	–	7	–0.97 (–3.80)***
%Outside Director Holdings	–	8	0.62 (1.06)
Firm Size	+ / –	Control Variable	–0.15 (–2.22)**
Adjusted R ²			0.24

***, **, * Significant at the 0.01, 0.05, and 0.10 levels, respectively.

^a t-statistics are after using the Froot (1989) procedure to adjust the variance-covariance matrix for dependence among observations from the same firm.

^b Variable definitions:

%Audout = $\ln(\%Audout / (1 - \%Audout) + 1)$;

Board Size = the natural log of the number of board members; and

%Outsiders = $\ln(\%Outsiders / (1 - \%Outsiders) + 1)$.

See Table 2 for other variable definitions.

I therefore re-estimate the regression in Table 4, substituting R&D Expenditures and a proxy, Hi-tech, for my growth opportunity measure. I define R&D Expenditures as research and development expenditures divided by total assets. Hi-tech is a dummy variable equal to 1 if the firm is in a high-tech industry, and 0 otherwise.¹³ Growth Opportunities is insignificantly positively correlated with R&D Expenditures ($\rho = 0.18$) and Hi-tech ($\rho = 0.13$).

¹³ Following Field and Hanka (2000), Hi-tech encompasses all firms with primary three-digit SIC codes in computer and office equipment (357), electronic components and accessories (367), miscellaneous electrical machinery, equipment, and supplies (369), laboratory apparatus and analytical, optical, measuring, and controlling instruments (382), surgical, medical, and dental instruments and supplies (384), and computer programming, data processing, and other computer-related services (737).

The coefficients on R&D Expenditures and Hi-tech are not significant in the multiple regression, suggesting that neither R&D Expenditures nor Hi-tech substitute for Growth Opportunities. I also re-estimate the regression in Table 4 after including Growth Opportunities along with R&D Expenditures and Hi-tech, respectively. The coefficients on Growth Opportunities exhibit virtually no change, and the coefficients on R&D Expenditures and Hi-tech remain insignificantly different from zero.

Accounting Losses and Legal Liability for Poor Performance

%Audout is lower for firms reporting losses in the previous two years. These results are consistent with the argument that investors in loss firms require less independent audit committees because financial information is less value-relevant for loss firms. A competing explanation for the negative association between losses and audit committee independence is that firms experiencing sustained losses have difficulty attracting outside directors to serve on their audit committees due to liability concerns. However, most large companies have director and officer insurance indemnifying non-inside directors from legal damages, and many states (especially Delaware and New York) severely limit lawsuits against non-inside directors.

Using the probability of bankruptcy as a proxy for legal liability, I substitute Altman's Z-statistic for Losses and also use Altman's Z-statistic as an additional regressor in the original multiple regression. In each regression, the coefficient on Altman's Z-statistic is insignificant at the 0.10 level, suggesting no systematic association with audit committee independence.

Alternative Definition of Audit Committee Independence

Current SEC and exchange rules suggest all members must be outside directors for the audit committee to be considered independent. As Table 1 shows, only 43.4 percent of sample firms had 100 percent independent audit committees in 1991–1993. To assess whether full independence is associated with the same economic factors as %Audout, I estimate a probit model using full independence (rather than %Audout) as the dependent variable. The inferences from the probit model are identical to those in Table 4, except the coefficient on CEO on Compensation Committee is significantly negative at the 0.05 level for the probit model.

Simultaneity

Some variables that explain %Audout also likely explain Board Size and %Outsiders. To account for simultaneities, I use a two-stage least squares (2SLS) method, in which %Audout is regressed on the factors used in Table 4, and either %Outsiders or Board Size is regressed on a set of endogenous and exogenous factors. The 2SLS estimator of %Outsiders or Board Size can be described as an instrumental variables estimator because this method substitutes instruments for %Outsiders or Board Size based on predicted values obtained from regressions of each variable on its set of factors.

Table 5 presents the 2SLS coefficients for each set of equations. The first two columns present the results with %Outsiders as the instrumental variable. To determine the simultaneous factors, I estimate stepwise regressions explaining %Outsiders and explaining Board Size, respectively. The set of possible explanatory variables include those used for the regression explaining %Audout and other potential determinants of board independence and board size described in the literature. I keep all explanatory variables from the stepwise regressions with p-values less than 0.15.

TABLE 5
Two-Stage Least Squares (2SLS) Regression of %Audout on %Outsiders, Board Size, and Other Endogenous and Exogenous Variables^a
 (coefficients and t-statistics)

	2SLS on %Audout and %Outsiders		2SLS on %Audout and Board Size	
	%Audout	%Outsiders	%Audout	Board Size
Intercept	1.09 (1.50)	1.41 (7.52)***	0.55 (0.92)	1.64 (28.52)***
%Outsiders	1.38 (3.70)***		1.84 (14.85)***	
Board Size	0.62 (2.35)**	-0.10 (-1.58)	0.67 (2.61)***	
Growth Opportunities	-0.08 (-2.31)**	-0.05 (-3.07)***	-0.05 (-1.88)*	
Losses	-0.54 (-1.91)*		-0.53 (-1.90)*	0.06 (1.94)*
Debt-to-Assets	-0.46 (-1.24)		-0.51 (-1.39)	
CEO on Compensation Committee	-0.62 (-2.95)***		-0.52 (-2.68)***	-0.06 (-2.28)**
5% Blockholder on Audit Committee	-1.24 (-4.82)***		-1.15 (-4.67)***	
%Outside Director Holdings	1.13 (1.63)	0.20 (0.92)	0.83 (1.27)	
Firm Size	-0.10 (-1.93)*		-0.12 (-2.17)**	
%Inside Director Holdings		-1.58 (-6.22)***		
CEO on Nominating Committee		-0.13 (-4.05)***		
CEO Tenure		-0.04 (-1.97)**		0.02 (2.14)**
R&D Expenditures		-0.88 (-2.03)**		-0.57 (-2.41)**
Beta		-0.13 (-3.18)***		-0.05 (-2.72)***
Institutional Ownership		0.41 (3.74)***		

(Continued on next page)

TABLE 5 (Continued)

	2SLS on %Audout and %Outsiders		2SLS on %Audout and Board Size	
	%Audout	%Outsiders	%Audout	Board Size
5% Outside Blockholder		0.21 (2.00)*		0.05 (1.16)
Lagged Stock Returns				0.08 (2.88)***
Hi-tech				-0.13 (-4.11)***
Adjusted R ²	0.10	0.16	0.28	0.37

***, **, * Significant at the 0.01, 0.05, and 0.10 levels, respectively.
a Variable definitions:
 %Inside Director Holdings = the percentage of shares owned by all inside directors;
 CEO on Nominating Committee = 1 if CEO sits on the board's nomination committee, and 0 otherwise;
 CEO Tenure = the natural log of the number of years the CEO has been on the board;
 R&D Expenditures = research and development expenses divided by total assets;
 Beta = the 120-day beta of the stock prior to the end of the last fiscal year;
 Institutional Ownership = the percentage of stock owned by institutions;
 5% Outside Blockholder = 1 if an outside director owns 5% or more shares, and 0 otherwise;
 Lagged Stock Returns = the one-year stock return prior to the shareholders' meeting; and
 Hi-tech = 1 if firm is in a high-technology industry, and 0 otherwise.
See Tables 2 and 4 for other variable definitions.

The endogenous variables for %Audout and %Outsiders are Board Size, Growth Opportunities, and %Outside Director Holdings. The exogenous variables for %Outsiders are %Inside Director Holdings (Hermalin and Weisbach 1988), CEO on Nominating Committee (Klein 1998), CEO Tenure (Hermalin and Weisbach 1988), R&D Expenditures (Reingold 1999), Beta, Institutional Ownership, and 5% Outside Blockholder. The last two columns present the results using Board Size as the instrumental variable. The endogenous variables are Losses and CEO on Compensation Committee. The exogenous variables for Board Size are CEO Tenure (Yermack 1996), R&D Expenditures (Reingold 1999), Beta, 5% Outside Blockholder, Lagged Stock Returns, and Hi-tech (Reingold 1999).

The inferences from the 2SLS analyses are similar to those derived from the multiple regression reported in Table 4. The one exception is that the coefficients on CEO on Compensation Committee are significantly negative at the 0.01 level for the 2SLS analyses, but insignificantly negative for the multiple regression. Thus, simultaneity does not affect the primary inferences drawn from Table 4, suggesting that the empirical results are robust to both procedures.

VI. CONCLUSIONS

Beginning in December 1999, the SEC and stock exchanges require listed firms to maintain audit committees with at least three directors, "all of whom have no relationship to the company that may interfere with the exercise of their independence from management and the company" (*NYSE Listed Company Manual* §303.01[B][2][a]). Although this statement suggests that firms must maintain audit committees composed solely of outside directors, exchange regulations allow for non-outside directors if the board determines it is in the firm's best interests for these individuals to serve on its audit committee.

I examine if variations in audit committee independence are associated with economic factors for a sample of S&P 500 firms over 1991–1993, a time period when firms had greater latitude in placing affiliated directors on their audit committee. I find that audit committee independence increases with board size and the percentage of outsiders on the board, consistent with the hypothesis that audit committee independence depends on the supply of available outside directors on the board. In contrast, audit committee independence decreases with the firm's growth opportunities and when the firm reported net losses in each of the two preceding years, supporting the hypothesis that audit committee independence is related to management's and shareholders' demand for scrutiny of the firm's financial accounting process. I also find a negative association between audit committee independence and the presence of alternative monitoring mechanisms, that is, for larger firms or when a nonmanagement director owning at least 5 percent of the firms' shares sits on the audit committee. Overall, my findings are consistent with the Blue Ribbon Commission's observation that "one size doesn't fit all" when it comes to audit committees. Thus, the stock exchanges may wish to allow boards of directors flexibility in determining their audit committee composition.

Several possible avenues for future research arise from this study. First, does the lower level of audit committee independence for higher growth firms and for firms with sustained losses result in higher incidences of financial fraud? Second, the Blue Ribbon Commission suggests that all audit committee members should have expertise in financial accounting. To what extent do audit committees comply with this suggestion, and what factors, if any, are related to fulfilling this mandate? Third, examining the interdependence between audit committee independence and competing corporate governance structures could further our understanding as to how corporations make trade-offs among these alternative mechanisms.

REFERENCES

- Amir, E., B. Lev, and T. Sougiannis. 1999. What value analysts? Working paper, Tel Aviv University, New York University, and University of Illinois at Urbana-Champaign.
- Barth, M. E., W. H. Beaver, and W. R. Landsman. 1998. Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting & Economics* 25: 1–34.
- Bhagat, S., and B. Black. 1999. The uncertain relationship between board composition and firm performance. *Business Lawyer* 54: 921–963.
- Carcello, J. V., and T. L. Neal. 2000. Audit committee composition and auditor reporting. *The Accounting Review* 75: 453–467.
- Collins, D. W., M. Pincus, and H. Xie. 1999. Equity valuation and negative earnings: The role of book value of equity. *The Accounting Review* 74: 29–61.
- DeFond, M. L., and J. Jiambalvo. 1994. Debt covenant violation and manipulation of accruals. *Journal of Accounting and Economics* 17: 145–176.
- Fama, E. F. 1980. Agency problems and the theory of the firm. *Journal of Political Economy* 88: 288–307.
- , and M. C. Jensen. 1983. Separation of ownership and control. *Journal of Law & Economics* 26: 301–325.
- Field, L. C., and G. Hanka. 2000. The expiration of IPO share lockups. *Journal of Finance* 56: 471–500.
- Froot, K. A. 1989. Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. *Journal of Financial and Quantitative Analysis* 24: 333–355.
- Hackman, J. R., ed. 1990. *Groups That Work and Those That Don't: Conditions for Effective Teamwork*. San Francisco, CA: Jossey-Bass.
- Hayn, C. 1995. The information content of losses. *Journal of Accounting & Economics* 20: 125–153.
- Hermalin, B. E., and M. S. Weisbach. 1988. The determinants of board composition. *RAND Journal of Economics* 19: 589–606.
- Jensen, M. C., and W. H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3: 305–360.
- . 1993. The modern industrial revolution, exit and the failure of internal control systems. *Journal of Finance* 48: 831–880.
- Klein, A. 1998. Firm performance and board committee structure. *Journal of Law and Economics* 41: 275–303.
- . 2001. Audit committees, board of director characteristics and earnings management. Working paper, New York University.
- Leftwich, R. 1983. Accounting information in private markets: Evidence from private lending agreements. *The Accounting Review*: 23–42.
- Lev, B., and P. Zarowin. 1999. The boundaries of financial reporting and how to extend them. *Journal of Accounting Research* 37: 353–385.
- Levitt, A. 1998. The numbers game. Remarks delivered at the NYU Center for Law and Business, New York, NY, September 28.
- Lipe, R. C., L. Bryant, and S. K. Widener. 1998. Do nonlinearity, firm-specific coefficients, and losses represent distinct factors in the relation between stock returns and accounting earnings? *Journal of Accounting & Economics* 25: 195–214.
- Lipton, M., and J. W. Lorsch. 1992. A modest proposal for improved corporate governance. *Business Lawyer* 48: 59–77.
- O'Reilly, V. M., P. J. McDonnell, B. N. Winograd, J. S. Gerson, and H. R. Jaenicke, eds. 1998. *Montgomery's Auditing*. 12th edition. New York, NY: John Wiley & Sons.
- The NASDAQ Stock Market Listing Requirements*. Undated. New York, NY: National Association of Securities Dealers.
- National Commission on Fraudulent Financial Reporting (the Treadway Commission). 1987. *Report of the National Commission on Fraudulent Financial Reporting*. Washington, D.C.: Government Printing Office.
- NYSE Listed Company Manual*. Undated. New York, NY: New York Stock Exchange.

- New York Stock Exchange and National Association of Securities Dealers. 1999. *Report and Recommendation of the Blue Ribbon Committee on Improving the Effectiveness of Corporate Audit Committees*. New York, NY: NYSE and NASD.
- Parker, S. 2000. The association between audit committee characteristics and the conservatism of financial reporting. Working paper, Santa Clara University.
- PricewaterhouseCoopers LLP. 2000. *Audit Committee Effectiveness—What Works Best*. 3rd edition. Altamonte Springs, FL: The Institute of Internal Auditors Research Foundation.
- Reingold, J. 1999. Dot.com boards are flouting the rules. *Business Week* (December 20): 130–134.
- Smith, C. W., and J. B. Warner. 1979. On financial contracting: An analysis of bond covenants. *Journal of Financial Economics* 7: 117–161.
- , and R. L. Watts. 1992. The investment opportunity set and corporate financing, dividend and compensation policies. *Journal of Financial Economics* 32: 263–292.
- Steiner, I. D. 1972. *Group Process and Productivity*. New York, NY: Academic Press.
- Verschoor, C. C. 1993. Benchmarking the audit committee. *Journal of Accountancy* (September): 59–64.
- Vicknair, D., K. Hickman, and K. C. Carnes. 1993. A note on audit committee independence: Evidence from the NYSE on “grey” area directors. *Accounting Horizons* 7 (March): 53–57.
- Watts, R. L., and J. L. Zimmerman. 1990. Towards a positive theory of positive accounting: A ten year perspective. *The Accounting Review* 65: 131–156.
- Williamson, O. E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*. New York, NY: The Free Press.
- Yermack, D. 1996. Higher market valuation of companies with a small board of directors. *Journal of Financial Economics* 40: 185–212.

Perspectives on Recent Capital Market Research

William H. Beaver
Stanford University

I. INTRODUCTION

The purpose of my remarks is to provide one perspective on major areas of capital markets research that have made important contributions to our understanding of accounting numbers, with an emphasis on those published during the past ten years. I do not intend these remarks to be a survey;¹ instead, I select the five research areas I believe have made the greatest contribution to our knowledge over the past ten years. These areas illustrate the degree to which capital market research has become interconnected. My remarks address why these areas are important, briefly summarize what we have learned, highlight some of the links between these areas, and raise some unresolved issues. Within each area, I identify major issues and some of the key papers, but I do not attempt to be comprehensive. In the process, I have sacrificed depth for breadth.

The five areas I have selected are market efficiency, Feltham-Ohlson modeling, value relevance, analysts' behavior, and discretionary behavior. The first two areas, market efficiency and Feltham-Ohlson modeling, are basic platforms that permit us to organize our thinking about the role of accounting in capital markets. The last three areas are applications that incorporate some form of accounting structure or individual behavior.

II. MARKET EFFICIENCY

Market efficiency is, of course, an important field of study. Much of the regulation of financial reporting is premised on the notion that once firms make accounting data publicly available, the implications will be widely appreciated and reflected in security prices. If the market is inefficient, then financial reporting and disclosure are not as effective, at least with respect to prices fully reflecting that information. Questions for regulators then arise as to whether altering the presentation of the data could mitigate this deficiency.

If investors trade in an efficient market, then they can rely on prices reflecting a rich set of the total mix of information, including financial statement information, and they need

¹ Kothari (2001) provides an excellent recent survey of capital markets research.

Editor's Note: The Executive Committee of the American Accounting Association has recommended the commissioning of a series of Presidential Research Lectures, to be delivered at annual meetings of the Association. To encourage broad dissemination, the Committee has requested that *The Accounting Review* publish this lecture given at the 2001 American Accounting Association Annual Meeting in Atlanta, GA.

Author's Note: I am pleased and honored to be the third Presidential Lecturer, following two distinguished colleagues, Shyam Sunder two years ago and Bill Kinney last year. I thank Mary Barth, Wayne Landsman, Maureen McNichols, and Stephen Ryan for many helpful suggestions on earlier drafts of this paper.

not process all of that information directly. In other words, investors become indirect beneficiaries of that information, even if they do not literally process it themselves. Efficient capital markets also have implications for resource allocation and production efficiency, so it is not surprising that market efficiency was one of the earliest areas studied.

Market efficiency is also of interest to researchers because, if they can assume market efficiency in the research design, then researchers can draw a different (and potentially more powerful) set of inferences. Assumptions about market efficiency affect the researchers' choice of the length of the window over which to compute abnormal returns in an event study. Market efficiency also affects the interpretations the researcher places on observed associations between security prices and accounting numbers.

There was a time when we thought that the issue of market efficiency with respect to publicly available data was "resolved." Early capital markets studies largely supported market efficiency. Both Ball and Brown (1968) and Beaver (1968) examined the post-announcement behavior of security returns and tentatively concluded that market efficiency was a reasonable approximation of the empirical results. Early studies examined changes in accounting methods (Archibald 1972; Ball 1972) and reached similar conclusions. Researchers also examined differences in accounting methods (e.g., Beaver and Dukes 1973) and concluded that the market prices behaved in a manner consistent with market efficiency.

Recent studies have reexamined market efficiency, and several have concluded that capital markets are inefficient with respect to at least three areas: post-earnings announcement drift, market-to-book ratios and its refinements, and contextual accounting issues.

Post-Earnings Announcement Drift

Post-earnings announcement drift was one of the first areas to suggest that markets may not be efficient with respect to accounting data. Several studies (Foster et al. 1984, among others) found evidence of post-earnings announcement drift in spite of attempts to control for at least some of the confounding factors. One reason for the finding of post-announcement drift is that the availability of daily return data enhances the power of the tests relative to prior research that used only weekly or monthly return data. The post-announcement drift studies culminated in Bernard and Thomas's (1989, 1990) research, which is an econometric tour de force. Their studies represent a classic example of excellent research design. The studies tenaciously pursue competing explanations, such as transaction costs vs. omitted risk factors, and find them lacking in many respects. Moreover, their studies, along with Freeman and Tse (1989), also explicitly develop an alternative hypothesis as to the nature of the market inefficiency and establish that the subsequent abnormal returns tend to cluster around subsequent earnings announcement dates. Their evidence that the abnormal returns are associated with some inefficient processing of earnings announcements is compelling.

An important extension is Abarbanell and Bernard (1992), whose study examines the question of whether a portion of the post-earnings announcement drift is attributable to the behavior of analysts' earnings forecasts and deficiencies in their processing of accounting data. This study reaches two important conclusions:

- (1) The analysts' forecasts appear to underestimate the persistency in earnings, and forecast errors based on analysts' forecasts are serially correlated. If analysts efficiently processed information, then the unconditional expected value of the forecast error would be zero, and the expected serial correlation would be zero.

- (2) This phenomenon explains some, but not all, of the post-earnings announcement drift. We would not expect analysts' behavior to be the complete explanation because post-earnings announcement drift is most pronounced in the small capitalization firms, which are not as heavily followed as larger capitalization firms are. Bartov et al. (2000) suggest that institutional holdings are also an important explanatory variable. In a further extension, Bhattacharya (2001) concludes that trade size, a proxy for less wealthy and less informed investors, may also be a factor. These studies represent a key link between market efficiency and the role of information and financial intermediaries.

Studies of post-earnings announcement drift are particularly compelling because earnings changes and earnings forecast errors have lower serial correlation than other candidates for market efficiency, such as market-to-book ratios. High serial correlation raises suspicion that the variable is a proxy for an omitted factor that is priced by the market.

Market-to-Book Ratios and Extensions

A second area of research is abnormal returns associated with portfolio strategies based on market-to-book ratios. Two early studies in this area are Fama and French (1992) and Lakonishok et al. (1994). The negative association between market-to-book ratios and subsequent returns appears to be significant and persistent and not explainable by conventional risk measures or their proxies. This is a controversial area; Fama (1991), among others, suggests that market-to-book ratios may represent some (otherwise unidentified) pricing factor, and Fama and French (1992) have posited a three-factor pricing model that includes the market-to-book ratio. As a result, recent market efficiency tests control for the market-to-book ratio and still find evidence of abnormal returns.

Major extensions of this work refine the market-to-book analysis by computing market-to-value ratios. At its simplest level, the argument for the market-to-book ratio as being a measure of market inefficiency is that the market prices either fail to reflect some factor related to the underlying value, which is reflected in book value, or that they include some factor unrelated to the underlying value. For example, high market-to-book ratio stocks represent so-called "glamour" stocks that are overpriced. From this perspective, one can extend the concept of value to include accounting "fundamentals" in addition to book value. Two major examples are Frankel and Lee (1998) and Dechow et al. (1999). Both studies employ modeling motivated by Feltham-Ohlson to predict the intrinsic value based on book value, earnings, and analysts' earnings forecasts. Both studies conclude that market-to-value ratios are associated with even higher subsequent abnormal returns than are the simpler market-to-book ratio strategies, which is consistent with these value estimates' being better proxies for underlying fundamentals and, hence, being better able to identify overpriced and underpriced stocks. As with research on post-earnings announcement drift, the abnormal returns here appear to be more prevalent in small capitalization stocks.

In another extension of the market-to-book research, Dechow and Sloan (1997) find that stock prices appear to reflect naively analysts' biased forecasts of future earnings growth, and reliance on analysts' earnings growth forecasts can explain over half of the higher returns associated with pursuing "contrarian" (e.g., market-to-book and price-earnings-based) strategies.

Contextual Accounting Issues

A key feature of the previous two areas is that they require little, if any, knowledge of the distinctive characteristics of how financial statements are prepared. They are based on

generic treatments of earnings and book value. Our comparative advantage as accounting researchers is in incorporating the richness of our knowledge of accounting institutions, reporting standards, and the composition of accounting numbers. Several recent studies have examined market efficiency based on some key feature of financial reporting.

Sloan (1996) is an excellent example of research that exploits our knowledge of one key feature: accrual accounting. Important aspects of the Sloan (1996) study are (1) an examination of the "consistency" between the weight placed on accruals and cash flow components in forecasting earnings and the implicit weight investors placed on the cash flow and accrual components of earnings in a valuation equation, and (2) the examination of portfolio strategies based on the magnitude of the accruals. Sloan (1996) concludes that capital markets overestimate the persistency of accruals and underestimate the persistency of cash flows from operations, because accruals are more subject to uncertainty of estimation and more subject to management and manipulation. Xie (2001) supports this conclusion by showing the mispricing documented by Sloan (1996) is largely due to abnormal accruals.

Moreover, in an attempt to address the "IPO puzzle," Teoh et al. (1998a, 1998b) and Teoh, Wong, et al. (1998) find that unusual accruals occur at Initial Public Offering dates and subsequently reverse. The accruals appear to be associated with at least a portion of the negative abnormal returns identified in the IPO research. These findings are consistent with those of Sloan (1996), as well as Xie (2001) and DeFond and Park (2001), that security prices do not fully reflect either the nature of accruals or their implications for future earnings and valuation.

However, not all accruals are associated with abnormal returns. Research also indicates that the supplemental disclosures with respect to specific accruals can permit capital markets to form unbiased estimates of the implications of the accrual for future earnings and hence for valuation. For example, Beaver and McNichols (2001) show that increased disclosure regarding the history of policy loss reserves in the property casualty insurance companies can make the accruals transparent to investors, and that revisions policy of loss accruals (development) are not associated with subsequent abnormal returns.

Unresolved Issues

The magnitude and length of the abnormal returns is surprising. For example, Frankel and Lee (1998) report that in the 36 months after portfolio formation, the abnormal returns associated with market-to-value strategies are 31 percent, whereas strategies that also exploit the predictability of analysts' forecasts are associated with abnormal returns of 45 percent.

There are several unresolved issues:

- (1) How can widely disseminated and examined data used with simple portfolio strategies that require no knowledge of accounting be associated with abnormal returns? From an economic perspective, widely disseminated data are not likely candidates.
- (2) How can studies of arcane disclosures (e.g., nonperforming loans and pensions, as in Beaver et al. [1989] and Barth et al. [1992]) find that such disclosures are apparently reflected in prices, yet more visible variables, such as earnings and book value, are not?
- (3) How can studies of security returns in the very short run (e.g., intraday returns, as in Patell and Wolfson [1984]) show evidence of relatively rapid response (within hours, if not minutes), and yet have evidence of abnormal returns that appear to persist for years after the portfolio formation date?

- (4) How can the body of research in aggregate show that prices both lead (e.g., Beaver et al. 1980; Beaver et al. 1987; Collins et al. 1997; Ryan 1995) and lag accounting data?

III. FELTHAM-OHLSON MODELING

For better or for worse, capital markets research is primarily empirical, rather than theoretical. One major exception is the modeling by Feltham and Ohlson (hereafter F-O). As one of the few attempts during the last ten years to develop a “theory of accounting” (i.e., a formal representation of value in terms of accounting numbers), the F-O approach is, in my opinion, one of the most important research developments in the last ten years (important articles are Ohlson [1995, 1999] and Feltham and Ohlson [1995, 1996]). F-O modeling is also one of the more controversial research areas in accounting. I will discuss the key features, empirical applications, major criticisms, and prospects for future research.

Key Features of F-O Modeling

One feature is the common set of assumptions that pervades the work. The assumptions include a valuation assumption that the value of equity is equal to the present value of expected future dividends, the clean surplus relation, and some form of a linear information dynamic. Feltham and Ohlson have derived a rich set of implications from these parsimonious assumptions.

In contrast to prior attempts to link accounting data and equity value, the F-O approach is neither a theory of information nor a theory of measurement. However, it permits a representation of the value of equity in terms of accounting numbers (most prominently, book value and expected abnormal earnings), relying essentially on the present value of expected dividends and clean surplus relation assumptions.

The F-O approach provides a role for many important features of the accounting system, including clean surplus, book value as well as earnings, transitory components of earnings, conservatism, and delayed recognition. For example, we can clearly see a progression in adding key features of the financial-reporting system. Early modeling assumed unbiased accounting and the clean surplus relation, while later extensions incorporated conservative accounting (Feltham and Ohlson 1995; Zhang 2000). Feltham-Ohlson alter the assumptions regarding the linear information dynamics to allow for “other information.” In doing so, the model provides a role for information that is currently known and reflected in price, but is reflected with a lag in the accounting numbers. The model provides a representation for delayed recognition. By enriching the linear information dynamics, Feltham and Ohlson (1996) construct a theory of depreciation, which distinguishes between two potential sources of conservatism: accelerated depreciation and positive net present value projects. Ohlson (1999) decomposes earnings into permanent and transitory components. In the process, he clearly distinguishes between the concepts of forecasting relevance and the time-series persistence of an earnings component, and he demonstrates how each relevance concept affects valuation relevance of that earnings component. Extensions alter the linear information assumptions to incorporate additional conditioning variables (e.g., different decompositions of earnings, such as cash flows and accruals, Barth et al. [1999]). Thus the F-O approach provides a potentially rich platform for further modifications of the linear information dynamics to address additional accounting issues of interest.

The F-O approach has stimulated considerable empirical research. Frankel and Lee (1998) and Dechow et al. (1999) use the approach in testing market efficiency, which is a very ambitious application of the model. Studies employing a combined book value and

earnings approach that either literally relies on the F-O model or is motivated by it are Barth, Beaver, and Landsman (1996, 1998), Barth and Clinch (1998), Burgstahler and Dichev (1997), Aboody et al. (1999), Barth, Beaver, Hand, and Landsman (1999), Collins, Maydew, and Weiss (1997), and Collins, Pincus, and Xie (1999). Other empirical applications include direct tests of the F-O model (e.g., Myers 1999; Barth, Beaver, Hand, and Landsman 1999). Major findings of this literature include the following:

- (1) Both book value and earnings are significant pricing factors.
- (2) The relative importance of book value is inversely related to the financial health of the firm.
- (3) The coefficient on earnings is lower for firms with low return on equity.
- (4) The coefficient on positive earnings is positive and significant, while the coefficient on losses is insignificantly different from zero.
- (5) Accrual vs. cash flow components of earnings are priced significantly differently from one another. In general, the accrual components are associated with a lower coefficient.

Criticisms of the Feltham-Ohlson Approach

One major criticism is that the model has no endogenous demand for accounting data, but how serious is this charge? The modeling can be informative without including an endogenous demand for accounting, and I believe the criticism is somewhat misplaced or misdirected. By analogy, the Capital Asset Pricing Model (CAPM) has no demand for financial institutions, yet we observe financial institutions empirically. What do we conclude? Do we conclude that the risk-return trade-off derived from the CAPM is of no interest or relevance to investors or to managers of financial institutions? I think not. The F-O models do not attempt to derive a demand for accounting. The F-O approach provides a framework for representing valuation in terms of accounting numbers, while taking accounting as given exogenously. This framework relates published accounting data to equity valuation, allows us to interpret the coefficients on the valuation equation, and allows us to relate the coefficients from the valuation equation to coefficients from the time-series of earnings equation. With contextual accounting arguments added to the general framework, researchers can predict how accounting numbers would relate to value (e.g., predictions on how the coefficients for the cash-flow and accrual components of earnings would be expected to differ in an earnings forecasting equation and a valuation equation).

Another criticism is that there is no information asymmetry, and that hence no strategic uses of accounting data arise within the F-O framework. To be sure, many financial-reporting issues arise out of concern over information asymmetry and incentives to "manage" the accounting numbers. For example, the research on analysts' behavior and discretionary behavior address issues of information asymmetry and incentives. However, the F-O approach is a beginning. As a prelude to developing models that incorporate information asymmetry and strategic uses of accounting data in valuation, it is helpful to start with a model of the relation between the valuation and accounting numbers in a nonstrategic setting. Moreover, not all issues of interest in accounting involve information asymmetry. A large body of research examines empirically the relation between valuation and publicly available accounting numbers in a nonstrategic setting (e.g., value-relevance studies). A conceptual framework, such as the F-O approach, guides the specification and interpretation of the empirical estimating equations.

Of course, it would also be desirable to have a theory where demand for accounting data is endogenous, and to have models of information asymmetry where incentives to

report strategically are endogenous. However, it is unreasonable to expect the F-O model to be rich enough to encompass all issues of interest to accounting research. Parsimony is a virtue in modeling. The model focuses attention on specific variables of interest, and the trade-off between insight and comprehensiveness is common in modeling exercises.

Some aspects of the models are unsupported by the empirical data (e.g., Myers 1999; Joos 2000; Barth et al. 1999), such as the linearity properties and the consistency among the coefficients in the system of linear information dynamics and valuation equations. However, the conflicting evidence highlights one of the important features of the F-O framework. Most accounting research is conducted in “reduced form.” In other words, we estimate intuitively plausible relations that we hope will allow us to predict the sign of the coefficient. Rarely do we predict the magnitude of the coefficient. Even more rare are opportunities to test structural relations among the coefficients across different equations in the system. The F-O models permit us to predict how the coefficients within and across equations in the system are related. In particular, the coefficients in the valuation equation are a function of the coefficients in the linear information dynamics equations.

Although it may seem disappointing that we can reject the null hypothesis that the predicted coefficients equal the empirically observed ones, it is progress to be in a position to specify a predicted relation among the coefficients. We make further progress by asking what modifications would permit a consistency between the systems of equations. Introducing nonlinearity in information dynamics is a likely candidate. For example, the F-O model does not incorporate bankruptcy or other option-related phenomena that might introduce nonlinearities into the relation. Research is currently underway to incorporate nonlinearities (Yee 2001).

The F-O modeling is one of the few attempts to pursue accounting theory. It is no coincidence that the terminology of accounting (e.g., *income*) is similar to that used in economics and finance. We could have called the difference between revenues and expenses by another name, but we did not. The semantics of accounting was chosen because income theory asserts that the resulting measure is an indicator of firm performance. Hence, it is natural to focus on a theory of measures of accounting net income and their relation to value.

Empirical studies applying the F-O framework often append a contextual accounting theory regarding differences in cash flows vs. accruals, the fair value of financial instruments, or the nature of pension obligations to guide the empirical predictions. Such contextual richness can help fill in some of the substance omitted from the parsimonious F-O representations. Once these contextual theories are appended, the combination of parsimonious modeling and contextual richness provides a rich basis for empirical testing. One of the major applications of the F-O models is the value-relevance literature.

IV. VALUE-RELEVANCE RESEARCH

Value relevance is major area of empirical research in the last ten years.² Holthausen and Watts (2001) identify 54 value-relevance studies, only three of which were published before 1990. Value-relevance research examines the association between a security price-based dependent variable and a set of accounting variables. An accounting number is termed “value relevant” if it is significantly related to the dependent variable.³ Defined in this most

² A more complete review of the literature appears in Barth, Beaver, and Landsman (2001).

³ Beaver (1998, 116), Ohlson (1995), and Barth (2000) provide closely related formal definitions. The key commonality is that an accounting amount is deemed value relevant if it is significantly associated with equity market value.

general sense, value-relevance research has a long history (Miller and Modigliani 1966); however, the term came into common usage in the early 1990s (Easton et al. 1993). Papers by Ohlson (1995, 1999) also use the term “value relevance,” in a manner consistent with empirical studies. As with the other research areas, value-relevance research is controversial (Holthausen and Watts 2001; Barth, Beaver, and Landsman 2001).

In this paper I address some basic questions: What distinguishes value-relevance research from other capital market research? Why is timeliness not a key issue in many value-relevance studies? What is the conceptual foundation of the value-relevance studies? What have we learned? What is the role of value-relevance research? What are some major unresolved issues?

What Are the Distinctive Characteristics?

Value-relevance research has two major characteristics. The first is that, more than any of the other four areas discussed, value-relevance research demands an in-depth knowledge of accounting institutions, accounting standards, and the specific features of the reported numbers. This knowledge includes the stated objectives of financial reporting, criteria standard setters use, the basis for specific standards, and details of how to construct the accounting numbers under a given standard (pension reporting is an excellent example). Incorporating the accounting context gives value-relevance research its richness and provides a basis for empirical predictions (Barth 1991), and vividly illustrates accounting researchers' comparative advantage in examining relations between equity value and accounting numbers.

A second distinguishing characteristic is that timeliness of information is not an overriding issue. Although value relevance research encompasses event studies, it also includes studies that examine the relation between the *levels* of stock prices and the accounting data. The timing of the information is of primary concern in the event-study research design. Event studies examine the stock price reaction over short windows of time centered on the announcement date. They identify the date of the public disclosure of the item being studied and examine the price change (usually in percentage terms and adjusted for market-wide movements) surrounding the event date.

In contrast to event studies, levels studies identify drivers of value that may be reflected in price over a longer time period than assumed in event studies. For example, prices may reflect the information before the announcement date. The value-relevance research characterizes market value at a point in time as a function of a set of accounting variables, such as assets, liabilities, revenues, expenses, and net income. This research design does not address timeliness, in contrast to “event-study” research design.

Why Is Timeliness Not the Key Issue?

The accounting system recognizes events later than security prices do (e.g., Ryan 1995). Delayed recognition is a natural implication of accounting standards, such as the revenue recognition principle. Moreover, we know that earnings announcements are largely, but not entirely, preempted by the disclosure of other information (Ball and Brown 1968; Beaver 1968). Landsman and Maydew (2002) conclude this finding has not changed over the last 30 years. Imagine a world in which earnings is the only information relevant to the value of the firm. With no private information search or prior public disclosures that preempt the earnings announcement, we would observe large spikes in price changes at earnings announcement times, in response to the unexpected earnings. However, this prospect creates incentives for private information search to obtain prior information about the forthcoming accounting earnings. To the extent that private information and prior public announcements

are reflected in prices before the public earnings announcement, the price reaction at the earnings announcement date is reduced. In the limit, a search for prior information can completely preempt the earnings announcement; however, such preemption does not eliminate the importance of reported earnings. The primary barrier to the complete preemption of earnings is the cost of obtaining the prior information. This cost includes not only the out-of-pocket cost of the information search, but also indirect costs imposed by the legal liability for selectively disseminating or obtaining nonpublicly available information. As the costs approach zero, the announcement effect can approach zero.

Models by Demski and Feltham (1994), McNichols and Trueman (1994), and Kim and Verrecchia (1994) formalize this process. For example, in Demski and Feltham (1994) the sole role of the information obtained before the earnings announcement is to provide information about forthcoming earnings. In short, these models imply there is more to the price-earnings relation than only the short-term price reactions at the announcement date. In fact, the magnitude of price change at the announcement date is informative about the costs of obtaining predisclosure information, but provides limited evidence about how value-relevant earnings are.

The informational approach states that a signal is informative only if the signal can alter beliefs conditional upon the other information available. This would require that the accounting number have some unique component that is not preempted by other information available prior to or simultaneous with the accounting number. This perspective is consistent with event study research designs, which control for other information publicly available prior to and concurrent with the accounting announcement.

However, accounting numbers are not unique representations of the underlying constructs they are designed to capture. It is often possible to find a vector of publicly available information that, collectively, is highly correlated with a particular accounting number. For example, the fair value of bank loans is a function of default risk and interest rate risk (Barth et al. 1996). Some linear combination of book value of the loans, proxies for default risk, and proxies for interest rate risk may be highly correlated with fair value measures, even if those measures “perfectly” capture the underlying construct. However, a key role of financial statements is to summarize relevant information parsimoniously and in a manner consistent with the underlying concept. It is informative to know how well accounting numbers play this role, even if vectors of competing proxies for the same underlying construct exist. In fact, if the accounting number (e.g., fair value of bank loans) is capturing the underlying construct, then we would expect other proxies for the construct (e.g., default risk and interest rate risk) to be correlated with the accounting number. Such correlation would indicate that the accounting number is capturing the underlying construct.

To choose another example, assume that an alternative set of data could produce a variable that is perfectly correlated with depreciation expense. Would this imply that one could exclude depreciation from the calculation of net income? Lambert (1996) concluded that the FASB probably would not exclude depreciation. The balance sheet and income statements are not intended to list only those assets, liabilities, revenues, and expenses not preempted by other publicly available information. The financial statements are intended instead to be “complete” within the constraints and definitions of generally accepted accounting principles. In this broader view of the role of financial statements, timeliness is only one dimension.

This broader view has implications for research design. For example, researchers often use first differences, rather than levels, of a stock-price-related dependent variable to mitigate some econometric problems, such as correlated omitted variables or serial dependency in the regression residuals (Landsman and Magliolo 1988). However, changing the form of

the variable may fundamentally change the question addressed. One chooses the levels design when the problem is to determine what accounting numbers are reflected in firm value, whereas one chooses the first differences research design when the problem is to explain changes in value over a specific period of time. Hence, in the first differences formulation, the issue of timing of the information is important. Thus, if the researcher is interested in whether the accounting amount is timely, then examining changes in value can be the appropriate research design choice. However, for the reasons discussed earlier, researchers are interested in a variety of questions, many of which do not involve timeliness.

What Is the Conceptual Foundation of Value-Relevance Research?

The theoretical foundation of value-relevance studies is a combination of a valuation theory plus contextual accounting arguments that allow researchers to predict how accounting variables relate to the market value of equity. There are three major types of valuation models. The oldest is an earnings-only approach—Miller and Modigliani (1966) characterize value as the present value of permanent future earnings. The research of Landsman (1986), Barth (1991), and Barth et al. (1996) adopts a balance-sheet approach. The Feltham-Ohlson model discussed earlier represents firm value as a linear function of book value of equity and the present value of expected future abnormal earnings. Value-relevance studies have relied heavily on a combined book value and earnings approach (Barth, Beaver, and Landsman 2001).

However, the valuation assumption is only half the story. Value-relevance studies typically incorporate contextual accounting arguments to predict the relation between accounting variables and market value. For example, the prediction that pension assets and obligations are priced as if they are assets and obligations of the company is based on the conceptual argument offered by the FASB (among others) regarding the economic substance of the pension contract between the company and employees, under a defined benefit plan (Landsman 1986).

The predictions of the way fair value of financial instruments will be priced in a valuation equation draws upon conceptual arguments concerning relevance and reliability of fair value vis-à-vis historical costs. This is an accounting theory, albeit one couched in terms of measurement of specific assets or obligations rather than a global statement that unequivocally predicts how all assets and liabilities would be measured and priced. Although the lack of a general theory of accounting can frustrate researchers (and others), researchers can use contextual accounting arguments to aid in predicting valuation-accounting number relations.

What Have We Learned?

The hallmark of value-relevance studies is that their execution requires an investment in and understanding of the institutional details of the way financial statements are prepared and of contextual arguments regarding the properties of various measures. What we have learned relates to three questions regarding an accounting number: Is it priced (i.e., does it have a coefficient that is significantly different from zero)? Is it priced consistently with some theoretical value (e.g., for a balance sheet number, is its coefficient equal to 1)? Is a particular accounting number priced equal to or differently from similar accounting numbers (e.g., do all components of net income have the same valuation multiples)? Here are some examples of what we have learned.

Evidence indicates that unrecorded pension assets and liabilities (unrecorded but disclosed in the footnotes) are priced. Landsman (1986), Barth (1991), and Barth et al. (1992) find that the unrecognized portion of pension assets and liabilities is priced in a manner

consistent with the capital market viewing pension assets as assets of the company and pension obligations as liabilities. Similar findings with respect to other post-retirement benefits (Amir 1993; Choi et al. 1997) are observed. Fair values of financial instruments are priced (Barth et al. 1996). However, the results are mixed with respect to the fair value of bank loans (Beaver and Venkatachalam 2000; Eccher et al. 1996; Nelson 1996). Moreover, Barth (1994a) shows that pricing multiples vary with the type of investment security in a manner related to the ease with which bank management can estimate the fair values.

Footnote information is often not as prominently displayed and may contain complex, arcane data (concerning such items as pension disclosures) that can be difficult to interpret. The pricing of footnote data is a nontrivial issue. The pricing of pension assets and obligations and the fair value of financial instruments is of interest in its own right. However, they are two prominent examples of the broader issue of whether footnote information is priced. Another example is nonperforming loans, which is footnote information on the default risk of bank loans. Empirical evidence indicates nonperforming loans are significant in explaining the value of bank common equity (Beaver et al. 1989; Wahlen 1994; Beaver and Engel 1996; Barth, Beaver, and Landsman 1996; Venkatachalam 1996).

The value-relevance literature also addresses questions relating to nonfinancial intangible assets. These studies generally find that intangible assets (e.g., capitalized software, brands, and goodwill) are priced (e.g., Aboody and Lev 1998; Barth, Clement, et al. 1998; Barth and Clinch 1998; Chambers et al. 1999). Studies also find that investors perceive research and development and advertising expenditures and bank core deposits as assets of the firm (e.g., Abdel-khalik 1975; Hirschey and Weygandt 1985; Bublitiz and Ettredge 1989; Landsman and Shapiro 1995; Barth et al. 1996; Eccher et al. 1996; Lev and Sougiannis 1996; Healy et al. 1997; Joos 2000). Barth and McNichols (1994) and Hughes (2000) find that unbooked environmental liabilities are also priced.

One would expect various components of earnings to be associated with different pricing multiples based on the persistence of that earnings component (Sloan 1996; Ohlson 1999). Empirical evidence indicates that the accrual components of earnings are not only less persistent than the cash-flow components in forecasting future earnings, but also that the accrual components are associated with a lower earnings multiple (Barth et al. 1999). Also for banks, earnings before security gains and losses is associated with a higher pricing multiple than security gains and losses (Barth et al. 1990).

The Role of Value-Relevance Research

Accounting research can play three roles:

- (1) Research can help articulate the nature of the issues, and can provide a paradigm or language with which to frame the questions of interest. The paradigm of the value of information is not a predictive theory in itself, but provides a definitional and taxonomic framework for formulating the informational role of accounting numbers.
- (2) Research can provide a theory. This theory can be normative, which leads to prescriptive statements, or positive, which provides hypotheses and testable predictions.
- (3) Research can provide empirical evidence.

Empirical evidence is a signal from an information system. The study's research design describes the features of the information system. The researcher forms priors with respect to the relationships of interest (e.g., probabilities that either of two alternative hypotheses is true). The evidence is a signal that leads to a posterior distribution, which must differ

from the prior for at least one possible signal for the research to be informative. As the power of the research design increases, the signals generated by the design become more informative.

Value-relevance research provides evidence as to whether the accounting numbers relate to value in the predicted manner. In the pension context, the predictions are based on contextual theory that pension assets are assets of the firm and pension obligations are obligations of the firm. A plausible prediction is that pension assets (obligations) are priced as assets (obligations). A study's findings represent only one of many possible outcomes. However, the subsequent discussions and the subsequent research conducted are informed by and conditioned on the observed evidence.

Unresolved Issues

Some of the unresolved issues affecting the inferences drawn from value-relevance research include market efficiency, econometric issues, and other purposes of financial statements. Does market efficiency affect interpretation of the results? I believe it does, but the findings are important even if markets are inefficient. Are the standard econometric issues more serious with respect to this research area than elsewhere? Most, if not all, of the econometric issues faced here are common to other areas of accounting research. Moreover, value-relevance research incorporates design features to mitigate these concerns. Finally, what other purposes of financial statements should be explored as a complement to the value relevance research? Accounting for contracting purposes is a major candidate (Watts and Zimmerman 1986).

V. RESEARCH ON ANALYSTS' BEHAVIOR

Another major research area is analysts' forecasting abilities and their coverage decisions. Analysts' behavior is important to accounting research, because analysts are among the major information intermediaries who use and interpret accounting data (Schipper 1991). As a result, security prices reflect the results of their analysis. Because the average prudent investor may lack the time, skill, or resources to analyze and interpret financial statements, analysts can be a major way in which accounting data become reflected in security prices. Efficient analysts' information processing can facilitate the efficiency of security prices, as well. If there are limitations and inefficiencies in the analysts' information processing, and if capital markets do not draw on other aspects of the total mix of information to circumvent analysts' limited information processing, then prices may not fully reflect the financial statement data. To the extent that analysts rely on a rich set of publicly available data, their forecasts can be a natural way to incorporate other information into the research design of valuation studies (e.g., via the application of the F-O models). An investigation of analysts' forecasts can assess the importance of accounting data relative to the total mix of information.

The history of analysts' forecasts is rich (Brown 1993). In some respects, it is the successor to the time-series of earnings literature (Beaver 1970; Ball and Watts 1972). The early literature focuses on which time-series model most accurately forecasts earnings. Identifying the process tells us something about the general characteristics of the accounting numbers (e.g., seasonality and adjacent quarter-to-quarter effects). Moreover, researchers use earnings forecasts derived from these models as inputs into other forms of research (e.g., we can use earnings forecast errors in security returns studies). Analysts' earnings forecasts are natural candidates for more accurate forecasts because they can reflect a richer information system than simply the past earnings series. One of the original purposes, learning about the features of the accounting system, has withered. However, the literature

has examined issues beyond those related to finding the most accurate earnings forecast. Biases, processing limitations, and strategic considerations have been addressed.

What Have We Learned?

Much prior research has concluded that analysts' forecasts are optimistic (O'Brien 1988, among others), although there appears to be secular reduction in the optimistic bias (Brown 2001). The degree of bias is related to underwriter affiliation. Analysts employed by investment firms that are associated with the underwriting of the firm's securities issue more optimistic forecasts (Lin and McNichols 1998). Analysts' (initially optimistic) forecasts tend to be revised downward during the year (Kasznik and McNichols 2001). Analysts with better forecasting ability appear to have a higher probability of survival (Mikhail, Walther, and Willis 1999; Clement 1999).

Analysts' forecasts outperform the best statistical models (Brown et al. 1987a, among others), which is not surprising, since the analysts can use a richer information set than the past earnings series. However, a model that incorporates both statistically based forecasts and analysts' forecasts outperforms analysts' forecasts alone, which implies that the analysts' forecasts do not reflect all of the information in the past earnings series (Brown et al. 1987b). The forecast errors based on analysts' forecasts are serially correlated, which is also consistent with the idea that analysts' forecasts do not fully reflect all the available information (Dechow et al. 1999; Frankel and Lee 1998). This evidence is also consistent with analysts' underestimating the persistency of earnings (Abarbanell and Bernard 1992).

These findings would be of mild interest in their own right even if capital markets fully adjusted for this behavior. However, they take on added significance, to the extent that capital markets do not appear to unravel these biases and processing inefficiencies. Capital markets appear to reflect naively analysts' forecasts in prices. This finding appears to explain (at least partially) the abnormal returns associated with market-to-book and market-to-value strategies (Dechow and Sloan 1997; Frankel and Lee 1998). Analysts' forecasts appear to be a parsimonious way to capture "other information" (at least in part) in the Ohlson sense of the term (Dechow et al. 1999). Analyst coverage is greater for firms with more institutional investors (O'Brien and Bhushan 1990) and more intangible assets (Barth, Kasznik, and McNichols 2001).

Unresolved Issues

Researchers need a better understanding of the incentives of analysts with respect to forecasting. In particular, why do analysts form biased forecasts? Even in the face of evidence that the bias is associated with underwriter affiliation, there are multiple explanations for the bias. Is it intentional, or is it a manifestation of self-selection (McNichols and O'Brien 1997)? Why do analysts misestimate the persistence of earnings? Why do forecasts not fully reflect the available information?

Do analysts learn over time? Are they more accurate with experience (Clement 1999)? Does the capital market learn over time in its processing of analysts' forecasts? How do analysts make decisions regarding the allocation of their efforts across the firms covered? How does analysts' behavior vary with the financial-reporting environment? For example, Barth, Kasznik, and McNichols (2001) find that analyst coverage increases with the presence of unrecorded intangible assets. What other financial-reporting features are important? Furthermore, what are the mechanisms by which analysts' forecasts are incorporated into price? Why do errors in analysts' forecasts appear to result in the mispricing of securities? Why does the market price appear not to adjust fully for these documented regularities in analysts' forecasts?

Another major issue is to identify the other information besides accounting data that influences analysts' forecasts. Frankel and Lee (1998) and Dechow et al. (1999) have used analysts' forecasts as a proxy for other information. However, from the context of a broader system, analysts' forecasts are endogenous and are a function of underlying exogenous variables. The dimensionality of such other information is quite large. However, it is important to identify at least some of the major exogenous variables that explain analysts' forecasts. Amir and Lev (1996), Deng et al. (1999), Ittner and Larcker (1998), Lev and Sougiannis (1996), and Joos (2000) explore the ability of nonfinancial measures, such as population within licensed areas, penetration ratios, patents, FDA approvals, concentration ratios, and market share to aid in predicting future earnings and in explaining prices. Is this information reflected in analysts' forecasts as well?

VI. RESEARCH ON DISCRETIONARY ACCRUALS

Management can improve or impair the quality of financial statements through the exercise of discretion over accounting numbers.⁴ Discretionary behavior includes voluntary earnings forecasting, voluntary disclosure, choice of accounting methods, and estimation of accruals. While research exists in all these areas, I will focus on the management of accruals (also known as earnings management). Accrual accounting is the heart of our financial-reporting system. I will discuss several aspects of earnings management: motives for earnings management, major findings, estimation of discretionary and nondiscretionary components, and unresolved issues.

Motives for Accrual Management

Motives fall into two broad categories: opportunistic or signaling. We tend to perceive the latter as benign, but not the former. Motives for managing accruals relate to compensation contracts, debt covenants, capital market pricing, taxes, litigation, and regulatory behavior (Watts and Zimmerman 1986; Beaver and Engel 1996). Each motive constitutes a broad category that encompasses a variety of specific behaviors. For example, capital market effects include management's attempts to influence the offering price in equity offerings, the terms or the value of stock options, and prices at which management-held securities are sold.

These motives can operate in either opposing or reinforcing ways, often making it difficult to isolate the primary motive (Healy and Wahlen 1999). For example, both capital market and compensation contracts can lead to incentives to overstate earnings. As a result, many researchers have not specified the precise nature of the underlying motivation, seeking instead to determine whether an empirical estimate of the discretionary accrual is related to some firm characteristic (e.g., financial difficulty, loss avoidance, income smoothing, big baths).

What Have We Learned?

Managers exercise discretion in response to a rich set of forces. Researchers use three major approaches to identify earnings management: generic models of discretionary accruals (e.g., Healy 1985; Jones 1991), tests based on discontinuities in the reported earnings distribution (e.g., Burgstahler and Dichev 1997), account-specific models of discretionary behavior (McNichols and Wilson 1988; Petroni 1992; Beatty et al. 1995), and combinations of these approaches (Beaver, McNichols et al. 2000).

⁴ McNichols (2000) reviews recent research in earnings management and discretion with respect to accounting data. Schipper (1989) provides an early perspective on earnings management.

Management appears to manage earnings to avoid a loss, to avoid an earnings decline (Burgstahler and Dichev 1997), and to avoid falling below analysts' forecasts (Burgstahler and Eames 1998). Firms that issue earnings forecasts tend to manage earnings toward meeting those forecasts (Kasznik 1999; Matsunaga and Park 2001). Earnings management appears to be widespread and relatively easy to detect, at least as estimated by extant techniques. Loan loss reserves in the banking sector and policy loss reserves in the insurance sector appear to be two major accounts subject to management. Accrual management is only one form of earnings management. Others include hedging activities (Barton 2001) and altering research and development expenditures (Bushee 1998). In the banking sector, management appears to manage the loan loss reserves jointly with other forms of earnings management (Beatty et al. 1995).

Capital markets appear to price differently the nondiscretionary and discretionary components of an accrual. In the banking sector, capital markets treat additional loan loss reserves as good news, not bad news, consistent with signaling interpretations of discretionary reporting of estimated loan losses. In particular, financially stronger banks signal they can afford to take the hit to earnings (Beaver et al. 1989; Wahlen 1994). Capital markets price discretionary components of loan loss reserves differently than nondiscretionary portions (Beaver and Engel 1996).

In the property and casualty sector, the development of policy loss reserves is consistent with earnings management (Beaver and McNichols 1998). The stock prices of property and casualty firms appear to reflect fully the predictability in the policy loss development (Beaver and McNichols 1998, 2001). However, the evidence from the generic accrual studies suggests a different picture with respect to market efficiency and the pricing of accruals. Accruals, considered to be more subject to discretion, are less persistent than stock prices of existing securities imply, whereas cash flow from operations is more persistent than stock prices imply (Sloan 1996; Xie 2001). Unusual accruals occur at initial public offering dates and reverse themselves subsequently. These unusual accruals appear to be correlated with the negative abnormal returns observed in the initial public offering literature (Teoh et al. 1998a, 1998b; Teoh, Wong et al. 1998).

Estimation of Discretionary and Nondiscretionary Accruals

A major issue with respect to the power of this research is the ability to identify proxies or conditioning variables that reflect the discretionary and nondiscretionary components of the accrual. In the Jones (1991) model, sales is the key nondiscretionary variable driving current accruals, and capital expenditures is the key variable driving noncurrent accruals. Needless to say, this is a parsimonious model. Research investigating sector-specific accruals, such as the loan loss provision, typically uses sector-specific variables, such as nonperforming loans, to increase the precision with which one can measure the nondiscretionary component.

Identifying proxies for discretionary accruals can be a challenge. Often, studies regress total accruals on only the nondiscretionary variables and assume the residual is discretionary (e.g., applications of the Jones model). Of course, failure to identify fully the nondiscretionary component implies the regression residual contains both discretionary and nondiscretionary components, and the researcher has measured the estimated discretionary and nondiscretionary components with error. Typically, the explicit conditioning variables for discretionary accruals, such as earnings or leverage, are generic. These generic variables can be proxies for many firm characteristics, which make the interpretation of coefficients of the discretionary accruals in valuation equations challenging.

The development of policy loss reserves in the property-casualty sector provides a unique opportunity to identify the discretionary component of an accrual (Petroni 1992).

Property-casualty firms must report *ex post* estimation error for reserves reported in earlier years, where the amount of this error is called development. Researchers can estimate the discretionary component without a specification of either discretionary or nondiscretionary variables. Development includes *ex post* surprises of a nondiscretionary nature. However, if development is not subject to discretion, then it has an expected value of zero, and, by implication, zero serial correlation. As discussed earlier, empirically, development has a positive expected value for financially weaker firms, implying an understatement of the loss reserve (Petroni 1992), and development is highly positively serially correlated over time, consistent with management recognizing information slowly over several years (Beaver and McNichols 1998). Petroni et al. (2000) further decompose the development into discretionary and nondiscretionary components and find they have different implications for future profitability, risk, and market value.

Unresolved Issues

Much of this discussion implies that extant methods for the identification of discretionary accruals are of potentially low power. Using an empirical simulation, Dechow et al. (1995) provide evidence that extant models are not very powerful. Yet, empirically, the majority of studies observe earnings management. Moreover, many forms of earnings management appear to be identifiable not only by researchers, but also by the capital markets.

Why is it relatively easy to detect earnings management empirically if the models are of low power? One might conjecture that effective earnings management (at least of the opportunistic type) would not be easy to unravel. Why is it relatively easy for the researchers to detect earnings management, typically using contemporaneous (not future) data? Is management naïve? Does earnings management achieve its goals (often unstated) even if it is invertible? What incentives for earnings management are consistent with the capital market's ability to invert the discretionary portion and price it differently? Is discretionary behavior a natural manifestation of contracting in incomplete markets (Demski and Frimor 1999)? The nature of the discretion may be known but not contractible. Incentives and costs to eliminate discretionary behavior are unclear, and discretionary behavior may be an equilibrium outcome, albeit not a "first best" solution.

These questions, of course, raise the possibility that what looks like earnings management may not be. Perhaps what researchers observe is not discretion or management at all, but is a proxy for some other factor. If so, then what might those factors be? McNichols (2000) offers evidence that discretionary accruals are correlated with growth and that the mispricing of accruals may in fact be the "glamour stock" phenomenon (i.e., the mispricing of high-expected-growth stocks) in disguise. In particular, she shows that aggregate accruals models that do not incorporate long-term earnings growth are potentially misspecified and can result in misleading inferences regarding earnings management. The implication is differential behavior observed in prior studies may relate to the performance characteristics of the firms (e.g., correlated with growth) rather than to differential incentives to manage earnings.

VII. CONCLUDING REMARKS

Market efficiency, Feltham-Ohlson modeling, value relevance, analysts' behavior, and discretionary behavior not only have had the greatest impact on capital market research over the last ten years, but they also have the greatest potential to contribute significantly to our knowledge over the next five to ten years. These areas address important questions. They are linked together and build upon one another's knowledge and research designs.

They raise major issues that remain unresolved. Three recurring themes are markets (efficiency, valuation), individual behavior (investors, analysts, managers), and accounting structure or context. Accounting research is distinct and important only insofar as it confronts the first two themes with the third—an observation similar in spirit to the point made by my predecessor (Kinney 2001).

Each of these research areas is controversial in terms of either findings or research method. Is important research by nature controversial? I believe controversy is a natural consequence of conducting important research, especially in the early stages of the research. Innovative research is likely to be the most controversial of all. Thirty-five years ago, many questioned whether capital market research with respect to accounting numbers was legitimate accounting research.

REFERENCES

- Abarbanell, J., and V. Bernard. 1992. Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *The Journal of Finance* 47 (July): 1181–1207.
- Abdel-khalik, A. R. 1975. Advertising effectiveness and accounting policy. *The Accounting Review* 50 (October): 657–670.
- Abodiy, D., and B. Lev. 1998. The value-relevance of intangibles: The case of software capitalization. *Journal of Accounting Research* 36 (Supplement): 161–191.
- , M. E. Barth, and R. Kasznik. 1999. Revaluations of fixed assets and future firm performance: Evidence from the U.K. *Journal of Accounting and Economics* 26 (January): 149–178.
- Amir, E. 1993. The market valuation of accounting information: The case of postretirement benefits other than pensions. *The Accounting Review* 68 (October): 703–724.
- , and B. Lev. 1996. Value-relevance of nonfinancial information: The wireless communication industry. *Journal of Accounting and Economics* 22 (1-3): 3–30.
- Archibald, T. 1972. Stock market reaction to depreciation switch-back. *The Accounting Review* 47 (January): 22–30.
- Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6 (Autumn): 159–178.
- . 1972. Changes in accounting techniques and stock prices. *Journal of Accounting Research, Empirical Research in Accounting: Selected Studies* (Supplement): 1–38.
- , and R. Watts. 1972. Some time series properties of accounting income. *The Journal of Finance* 27 (June): 663–681.
- Barth, M. E., W. H. Beaver, and M. Wolfson. 1990. Components of earnings and the structure of bank share prices. *Financial Analysts Journal* 46 (May/June): 53–60.
- . 1991. Relative measurement errors among alternative pension asset and liability measures. *The Accounting Review* 66 (July): 433–463.
- , W. H. Beaver, and W. R. Landsman. 1992. The market valuation implications of net periodic pension cost components. *Journal of Accounting and Economics* 15 (March): 27–62.
- . 1994a. Fair value accounting: Evidence from investment securities and the market valuation of banks. *The Accounting Review* 69 (January): 1–25.
- . 1994b. Fair-value accounting for banks' investment securities: What do bank share prices tell us? *Bank Accounting and Finance* 7 (Summer): 13–23.
- , and M. McNichols. 1994. Estimation and market valuation of environmental liabilities relating to superfund sites. *Journal of Accounting Research* 32 (Supplement): 177–209.
- , W. H. Beaver, and W. R. Landsman. 1996. Value-relevance of banks' fair value disclosures under SFAS 107. *The Accounting Review* 71 (October): 513–537.
- , ———, and ———. 1998. Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting and Economics* 25 (February): 1–34.
- , M. B. Clement, G. Foster, and R. Kasznik. 1998. Brand values and capital market valuation. *Review of Accounting Studies* 3: 41–68.

- , and G. Clinch. 1998. Revalued financial, tangible, and intangible assets: Associations with share prices and non market-based value estimates. *Journal of Accounting Research* 36 (Supplement): 199–233.
- , W. H. Beaver, J. R. M. Hand, and W. R. Landsman. 1999. Accruals, cash flows, and equity values. *Review of Accounting Studies* 4 (December): 205–229.
- . 2000. Valuation-based research implications for financial reporting and opportunities for future research. *Accounting and Finance* 40: 7–31.
- , W. H. Beaver, and W. R. Landsman. 2001. The relevance of value relevance research for financial accounting standard setting: Another view. *Journal of Accounting and Economics* 39 (September): 77–104.
- , R. Kasznik, and M. McNichols. 2001. Analyst coverage and intangible assets. *Journal of Accounting Research* 39 (June): 1–34.
- Barton, J. 2001. Does the use of financial derivatives affect earnings management decisions? *The Accounting Review* 76 (January): 1–26.
- Bartov, E., S. Radhakrishnan, and I. Krinsky. 2000. Investor sophistication and patterns in stock returns after earnings announcements. *The Accounting Review* 75 (January): 43–63.
- Beatty, A., S. Chamberlain, and J. Magliolo. 1995. Managing financial reports of commercial banks: The influence of taxes, regulatory capital and earnings. *Journal of Accounting Research* 33 (Autumn): 231–261.
- Beaver, W. H. 1968. The information content of earnings. *Journal of Accounting Research* 6 (Supplement): 67–92.
- . 1970. The time series behavior of earnings. *Journal of Accounting Research* 8 (Supplement): 62–99.
- , and R. Dukes. 1973. Interperiod tax allocation and delta depreciation methods: Some empirical results. *The Accounting Review* 48 (July): 549–559.
- , R. Lambert, and D. Morse. 1980. The information content of security prices. *Journal of Accounting and Economics* 2 (March): 3–28.
- , ———, and S. Ryan. 1987. The information content of security prices: A second look. *Journal of Accounting and Economics* 9 (July): 139–157.
- , C. Eger, S. Ryan, and M. Wolfson. 1989. Financial reporting, supplemental disclosure, and bank share prices. *Journal of Accounting Research* 27 (Autumn): 157–178.
- , and E. Engel. 1996. Discretionary behavior with respect to allowances for loan losses and the behavior of security prices. *Journal of Accounting and Economics* 22 (August–December): 177–206.
- . 1998. *Financial Reporting: An Accounting Revolution*. 3rd edition. Englewood Cliffs, NJ: Prentice Hall.
- , and M. F. McNichols. 1998. The characteristics and valuation of loss reserves of property casualty insurers. *Review of Accounting Studies* 3: 73–95.
- , ———, and K. Nelson. 2000. Management of loss reserve accrual and the distribution of earnings in the property-casualty insurance industry. Working paper, Stanford University.
- , and M. Venkatachalam. 2000. Differential pricing of the discretionary and nondiscretionary components of loan fair values. Working paper, Stanford University.
- , and M. F. McNichols. 2001. Do stock prices of property casualty insurers fully reflect information about earnings, accruals, cash flows and development? *Review of Accounting Studies* 6 (June): 197–220.
- Bernard, V. L., and J. Thomas. 1989. Post-earnings announcement drift: delayed price response or risk premium? *Journal of Accounting Research* 27 (Supplement): 1–48.
- , and ———. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13 (December): 305–340.
- Bhattacharya, N. 2001. Investors' trade size and trading responses around earnings announcements: An empirical investigation. *The Accounting Review* 76 (April): 221–244.

- Brown, L., P. Griffin, R. Hagerman, and M. Zmijewski. 1987a. Security analyst superiority relative to univariate time-series models in forecasting of quarterly earnings. *Journal of Accounting and Economics* 9 (April): 61–88.
- , ———, ———, and ———. 1987b. An evaluation of alternative proxies for the market's assessment of unexpected earnings, *Journal of Accounting and Economics* 9 (July): 159–193.
- . 1993. Earnings forecasting research: Its implications for capital market research. *International Journal of Forecasting* 9: 295–320.
- . 2001. A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research* 39 (September): 221–241.
- Bublitz, B., and M. Ettredge. 1989. The information in discretionary outlays: Advertising, research and development. *The Accounting Review* 64 (January): 108–124.
- Burgstahler, D. C., and I. D. Dichev. 1997. Earnings, adaptation and equity value. *The Accounting Review* 72 (April): 187–215.
- , and M. Eames. 1998. Management of earnings and analysts forecasts. Working paper, University of Washington, Seattle.
- Bushee, B. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73 (July): 305–333.
- Chambers, D., R. Jennings, and R. B. Thompson III. 1999. Evidence on the usefulness of capital expenditures as an alternative measure of depreciation. *Review of Accounting Studies* 4 (December): 169–195.
- Choi, B., D. W. Collins, and W. B. Johnson. 1997. Valuation implications of reliability differences: The case of nonpension postretirement obligations. *The Accounting Review* 72 (July): 351–383.
- Clement, M. 1999. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics* 27 (June): 285–303.
- Collins, D. W., E. L. Maydew, and I. S. Wiess. 1997. Changes in the value-relevance of earnings and equity book values over the past forty years. *Journal of Accounting and Economics* 24 (December): 39–67.
- , M. Pincus, and H. Xie. 1999. Equity valuation and negative earnings: The role of negative earnings. *The Accounting Review* 74 (January): 29–61.
- Dechow, P. M., R. G. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (April): 193–225.
- , and ———. 1997. Returns to contrarian investment strategies: Tests of naïve expectations hypotheses. *Journal of Financial Economics* 43 (January): 3–27.
- , A. P. Hutton, and R. G. Sloan. 1999. An empirical assessment of the residual income valuation model. *Journal of Accounting and Economics* 26 (January): 1–34.
- DeFond, M. and C. Park. 2001. The reversal of abnormal accruals and the market valuation of earnings surprises. *The Accounting Review* 76 (July): 375–404.
- Demski, J., and G. Feltham. 1994. Market response to financial reports. *Journal of Accounting and Economics* 17 (January): 3–40.
- , and R. Frimor. 1999. Performance garbling under renegotiation in multi-period agencies. *Journal of Accounting Research* 37 (Supplement): 187–214.
- Deng, Z., B. Lev, and F. Narin. 1999. Science and technology as predictors of stock performance. *Financial Analysts Journal* 55 (May/June): 20–32.
- Easton, P. D., P. H. Edey, and T. S. Harris. 1993. An investigation of revaluations of tangible long-lived assets. *Journal of Accounting Research* 31 (Supplement): 1–38.
- Eccher, E. A., K. Ramesh, and S. R. Thiagarajan. 1996. Fair value disclosures by bank holding companies. *Journal of Accounting and Economics* 22 (August–December): 79–117.
- Fama, E. 1991. Efficient capital markets II. *The Journal of Finance* 46 (December): 1575–1617.
- , and K. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47 (July): 427–465.
- Feltham, G. A., and J. A. Ohlson. 1995. Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research* 11 (Spring): 689–732.

- , and ———. 1996. Uncertainty resolution and the theory of depreciation measurement. *Journal of Accounting Research* 34 (Autumn): 209–234.
- Foster, G., C. Olsen, and T. Shevlin. 1984. Earnings releases, anomalies and the behavior of security returns. *The Accounting Review* 59 (October): 574–603.
- Frankel, R. M., and C. M. C. Lee. 1998. Accounting valuation, market expectation, and the cross-sectional stock returns. *Journal of Accounting and Economics* 25 (June): 283–319.
- Freeman, R., and S. Tse. 1989. The multiperiod information content of accounting earnings: Confirmations and contradictions of previous earnings reports. *Journal of Accounting Research* 27 (Supplement): 49–79.
- Healy, P. M. 1985. The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics* 7 (April): 85–107.
- , S. Myers, and C. Howe. 1997. R&D accounting and the relevance-objectivity trade-off: A simulation using data from the pharmaceutical industry. Working paper, Harvard Business School.
- , and J. Wahlen. 1999. A review of the earnings management literature and its implications for standard setting. *Accounting Horizons* 13 (December): 365–383.
- Hirschey, M., and J. J. Weygandt. 1985. Amortization policy for advertising and research and development expenditures. *Journal of Accounting Research* 23 (Spring): 326–335.
- Holthausen, R. W., and R. L. Watts. 2001. The relevance of the value relevance literature for financial accounting standard setting. *Journal of Accounting and Economics* 31 (September): 3–75.
- Hughes, K. 2000. The value relevance of nonfinancial measures of air pollution in the electric utility industry. *The Accounting Review* 75 (April): 209–228.
- Ittner, C., and D. Larcker. 1998. Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research* 36 (Supplement): 1–35.
- Jones, J. J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29 (Autumn): 193–228.
- Joos, P. 2000. Explaining cross-sectional differences in unrecognized net assets. Working paper, Stanford University, June.
- Kasznik, R. 1999. On the association between voluntary disclosure and earnings management. *Journal of Accounting Research* 37 (Spring): 57–81.
- , and M. McNichols. 2001. Does meeting expectations matter? Evidence from analyst forecast revisions and share prices. Working paper, Stanford University.
- Kim, O., and R. Verrecchia. 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17 (January): 41–67.
- Kinney, W. 2001. Accounting scholarship: What is uniquely ours? *The Accounting Review* 76 (April): 275–284.
- Kothari, S. P. 2001. Capital markets research in accounting. *Journal of Accounting and Economics* 31 (September): 105–231.
- Lakonishok, J., A. Shleifer, and R. Vishny. 1994. Contrarian investment, extrapolation, and risk. *Journal of Finance* 49 (December): 1541–1578.
- Lambert, R. 1996. Financial reporting research and standard setting. Working paper, Stanford University.
- Landsman, W. 1986. An empirical investigation of pension fund property rights. *The Accounting Review* 61 (October): 662–691.
- , and J. Magliolo. 1988. Cross-sectional capital market research and model specification. *The Accounting Review* 63 (October): 586–604.
- , and A. L. Shapiro. 1995. Tobin's Q and the relation between accounting ROI and economic return. *Journal of Accounting, Auditing and Finance* 10 (Spring): 103–121.
- , and E. Maydew. 2002. Beaver (1968) revisited: Has the information content of quarterly earnings announcements declined in the past three decades? *Journal of Accounting Research* (forthcoming).
- Lev, B., and T. Sougiannis. 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics* 21 (February): 107–138.

- Lin, H., and M. McNichols. 1998. Underwriting relationships and analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics* 25 (February): 101–127.
- Matsunaga, S., and C. Park. 2001. The effect of missing a quarterly earnings benchmark on the CEO's annual bonus. *The Accounting Review* 76 (July): 313–332.
- McNichols, M., and G. P. Wilson. 1988. Evidence of earnings management from the provision for bad debts. *Journal of Accounting Research* 26 (Supplement): 1–31.
- , and B. Trueman. 1994. Public disclosure, private information collection, and short term trading. *Journal of Accounting and Economics* 17 (January): 69–94.
- , and P. O'Brien. 1997. Self-selection and analyst coverage. *Journal of Accounting Research* 35 (Supplement): 167–199.
- . 2000. Research design issues in earnings management studies. *Journal of Accounting and Public Policy* 19 (Winter): 313–345.
- Mikhail, M., B. Walther, and R. Willis. 1999. Does forecast accuracy matter to security analysts? *The Accounting Review* 74 (April): 185–200.
- Miller, M. H., and F. Modigliani. 1966. Some estimates of the cost of capital to the electric utility industry, 1954–57. *The American Economic Review* 56 (June): 333–391.
- Myers, J. 1999. Implementing residual income valuation with linear information dynamics. *The Accounting Review* 74 (January): 1–28.
- Nelson, K. 1996. Fair value accounting for commercial banks: An empirical analysis of SFAS No. 107. *The Accounting Review* 71 (April): 161–182.
- O'Brien, P. 1988. Analysts' forecasts as earnings expectations. *Journal of Accounting and Economics* 10 (January): 53–83.
- , and R. Bhushan. 1990. Analyst following and institutional ownership. *Journal of Accounting Research* 28 (Supplement): 55–76.
- Ohlson, J. 1995. Earnings, book values and dividends in security valuation. *Contemporary Accounting Research* 11 (Spring): 661–687.
- . 1999. On transitory earnings. *Review of Accounting Studies* 4 (December): 145–162.
- Patell, J., and M. Wolfson. 1984. The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of Financial Economics* 13 (June): 223–252.
- Petroni, K., 1992. Optimistic reporting in the property-casualty insurance industry. *Journal of Accounting and Economics* 15 (December): 485–508.
- , S. Ryan, and J. Wahlen. 2000. Discretionary and nondiscretionary revisions of loss reserves by property casualty insurers: Differential implications for future profitability, risk and market value. *Review of Accounting Studies* 15 (June): 95–125.
- Ryan, S. 1995. A model of accrual measurement with implications for the evolution of the book-to-market ratio. *Journal of Accounting Research* 33 (Spring): 95–112.
- Schipper, K. 1989. Commentary on earnings management. *Accounting Horizons* 3 (December): 91–102.
- . 1991. Commentary on analysts' forecasts. *Accounting Horizons* 5 (December): 105–121.
- Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (July): 289–315.
- Teoh, S. H., I. Welch, and T. J. Wong. 1998a. Earnings management and the long-run market performance of initial public offerings. *The Journal of Finance* 53 (December): 1935–1974.
- , ———, and ———. 1998b. Earnings management and the post-issue underperformance of seasoned public offerings. *Journal of Financial Economics* 50 (October): 63–99.
- , T. J. Wong, and G. Rao. 1998. Are accruals during initial public offerings opportunistic? *Review of Accounting Studies* 3: 175–208.
- Venkatachalam, M. 1996. Value-relevance of banks' derivatives disclosures. *Journal of Accounting and Economics* 22 (August-December): 327–355.
- Wahlen, J. 1994. The nature of information in commercial bank loan loss disclosures. *The Accounting Review* 69 (July): 455–478.
- Watts, R. L., and J. L. Zimmerman. 1986. *Positive Accounting Theory*. Engelwood Cliffs, NJ: Prentice Hall.

- Xie, H. 2001. The mispricing of abnormal accruals. *The Accounting Review* 76 (July): 357–373.
- Yee, K. 2001. Modigliani-Miller invariant valuation functions. Working paper, Stanford University.
- Zhang, X. 2000. Conservative accounting and equity valuation. *Journal of Accounting and Economics* 29 (February): 125–149.

BOOK REVIEWS

Jane L. Reimers, Editor

Editor's Note: The book reviews in this issue address very current topics in accounting, and they provide ideas and insights useful for both teaching and research. Thanks to both of the reviewers for their comments on these two recently published books.

SUNSTEIN, CASS R., ed. *Behavioral Law and Economics* (Cambridge University Press, New York, NY, 2000, 421 pp.)

Behavioral Law and Economics is a new addition to the Cambridge University Press Series on Judgment and Decision Making. Like others in the series, this book is a well-selected set of articles and essays. Most have been rewritten for this book to make the text coherent. Authors of chapters include prominent scholars in psychology (e.g., Daniel Kahneman, the late Amos Tversky, David Schkade, and George Loewenstein) and behavioral economics (e.g., Richard Thaler and Jack Knetsch) who are familiar to accounting academicians.

The theme of this volume, like others in the series, is the descriptive validity of behavioral science, particularly psychology. Constraints on human information processing are (1) the number of things one can keep in mind at the same time are few, (2) thinking is a serial process, and (3) information in memory, while apparently limitless in quantity, is not accessible error-free (to put it mildly). The information-processing capabilities required for economically rational choice exceed these limitations (e.g., no instance of complete and transitive preference has been documented). Even though economic rationality conforms to the soundest of syllogistic logic, it is descriptively untrue of human judgment.

For example, a common strategy is the so-called framing of risky choices. Individuals view possible outcomes through frames of reference. The payoff of each option is *measured* relative to a reference point. A prize that in absolute terms brings a \$5 inflow, could be measured at $-\$10$ (if the reference point were \$15) or plus \$20 (if the reference point were at $-\$15$).

The framing compatible with rational choice is ending wealth position in which the payoff from each option is integrated with current wealth, and then the best option is selected by comparing the ending wealth position given each option. The framing of outcomes around the amount at issue simplifies the comparison of the options relative to comparisons of ending portfolios.

Though strategies such as framing may result in choices that are inconsistent relative to a "reasonable person" standard in law (or an optimizing choice by the standard of the economic decision model), they are reasonable in that they conserve the limited and scarce cognitive capacities available for problem solving. That is, human information *processing* is not a free good.

This volume, like others in the series, is in no way negative or sensationalist. It applies this behavioral perspective to law and economics. The theme is that patterns of institutional practice in law and economics are the product of human information-processing limitations. The way humans bring the cognitive processing demands of a judgment within limits of processing capacity gives structure and content to our legal and economic institutions. Features of the decision context (e.g., extrinsic incentives involved) and individual differences (e.g., expertise) are important.

To continue the example of framing, taxpayers prefer a tax collected such that they receive only the after-tax amount to a tax system in which they receive a gross amount and then must pay the tax (experiencing a loss), even if the two options are economically equivalent. Thus, tax payment through withholding and corporate income taxes are popularly preferred to requiring taxpayers to pay the taxes directly out of total income. Also, in labor

negotiations an offer of a new benefit with a fixed value (a potential gain) would not be as valuable as retaining a current benefit of the same value (a potential loss).

This book should be of interest to accounting researchers. Tax accounting, auditing, financial reporting, and cost accounting measures involve legal rights and obligations created in legislative, administrative, and contractual law. Thus, the subject matter of behavioral law is directly applicable to accounting.

One of the four sections is "The Demand for Law." Analogous to the call for understanding the economic demand for accounting from the Rochester School some 30 years ago, this section suggests that behavioral factors explain why law is as it is. For another example, one essay argues that an optimism that is helpful to the functioning of large organizations induces an environment in which managers must hide bad news. The implication for accounting research of earnings management is obvious.

Many of the behavioral features of law described here are applicable to accounting. The book is readable by persons largely unfamiliar with the extensive behavioral research that supports the analyses, including practitioners as well as accounting researchers whose work is economics-based. Accounting researchers whose work is psychology-based will find the applications intriguing.

RICHARD B. DUSENBURY
Associate Professor

Florida State University

GLOVER, STEVEN M., STEPHEN W. LIDDLE, and DOUGLAS F. PRAWITT, *E-Business: Principles and Strategies for Accountants* (Upper Saddle River, NJ: Prentice Hall, 2001, 170 pp.)

This book is intended to provide a relatively nontechnical introduction to the world of e-business—a world that accounting professionals in training must be exposed to. The authors do an outstanding job of presenting the material in a way that is accessible to a broad audience and is well suited as a supplement to most accounting courses at the undergraduate level. While the authors identify intermediate and advanced financial accounting courses, introductory or advanced auditing courses, and introductory systems courses as the likely market, I would also consider the inclusion of the material at the principles of accounting level as an ancillary source to provide an early overview of e-business and the accounting function to be appropriate.

The first three chapters are especially well suited to a broad audience. Chapter 1 defines e-business and provides a brief overview of benefits and risks introduced by e-business enabling technology. The chapter also identifies opportunities and challenges that e-business presents to the accounting profession. The second chapter presents a relatively brief explication of the development of the Internet and electronic data interchange and provides a useful glossary of e-business terminology (supplemented by several more technical appendices). Chapter 3 contains an excellent discussion of various business models associated with e-business (e.g., business-to-consumer, business-to-business) and follows up with several two-to-three page overviews of specific businesses that have successfully implemented e-business strategies. Four companies (Dell, Amazon, eBay, and Charles Schwab) are highlighted and should prove interesting to students.

Throughout the book, additional sidebars and highlighted text boxes provide additional richness to the presentation by providing brief vignettes or expanded discussions of points introduced in the main body of the text. Importantly, the text is supported by a well-designed web site (<http://www.prenhall.com/glover>) that will prove useful to adopting instructors as well as their students. The web site is also organized by chapter to correspond with the book and provides additional and updated sources of information to complement the material presented in each chapter. At the end of every chapter, questions and cases are provided to stimulate classroom discussion and to provide opportunities for instructors to evaluate student comprehension.

Chapters 4 and 5 are more narrowly focused and would, perhaps, prove challenging to an introductory class. Chapter 4, "Identifying and Managing the Risks of E-Business," expands significantly on the brief discussion in the first chapter. Specific audit risks central to a computerized environment are identified along with controls that may be implemented to mitigate such risks. Some knowledge of the computing environment is assumed, and thus this discussion is probably best targeted toward students in a junior-level accounting information systems course or as a component of a senior-level auditing course.

Chapter 5, "Implications and Opportunities of e-Business Assurance," focuses on specific features of e-business models that present risk exposures to the accounting and auditing function. For example, there is an insightful discussion of revenue recognition issues that are unique to the e-business setting. Given the level of attention that has been directed to this issue by the Securities and Exchange Commission in the recent past, this provides an excellent opportunity to expose students to the SEC's role in financial reporting. This chapter also includes an overview of assurance services that are being developed to specifically address e-business needs. CPA

WebTrust, SysTrust, and Verisign are all presented as examples of e-assurance products that have been created to fill voids that have developed in response to emerging business models and practices.

As noted earlier, the book includes technical appendices (six in all) and also a useful index. I believe this text could be used to introduce business students in all concentrations to e-business concepts germane to their future professional careers. Accounting majors in particular would profit from an early exposure to these topics. The authors are to be commended for tackling a technical subject with grace and enthusiasm—the text reads easily and contains enough “real-world” examples to hold most readers attention.

BRUCE C. BRANSON
Associate Professor

North Carolina State University

Editorial Data

The following table contains information about turnaround time for manuscripts (including revisions) on which editorial decisions were made in the 12-month period ended November 30, 2001. Turnaround time is the number of days between the date that the manuscript was received *and the date of the editor's letter to the author(s)*.

	Number of Manuscripts	Cumulative Number	Cumulative Percent
$0 \leq \text{Days} \leq 30$	82	82	27.61
$31 \leq \text{Days} \leq 60$	114	196	65.99
$61 \leq \text{Days} \leq 90$	65	261	87.88
$91 \leq \text{Days} \leq 120$	28	289	97.31
$121 \leq \text{Days}$	8	297	100.00

The mean review time was 51 days; the median review time was 50 days.

EDITORIAL POLICY AND STYLE INFORMATION

EDITORIAL POLICY

According to the policies set by the Publications Committee (which were endorsed by the Executive Committee and were published in the *Accounting Education News*, June 1987), *The Accounting Review* "should be viewed as the premier journal for publishing articles reporting the results of accounting research and explaining and illustrating related research methodology. The scope of acceptable articles should embrace any research methodology and any accounting-related subject, as long as the articles meet the standards established for publication in the journal....No special sections should be necessary. The primary, but not exclusive, audience should be—as it is now—academicians, graduate students, and others interested in accounting research."

The primary criterion for publication in *The Accounting Review* is the significance of the contribution an article makes to the literature.

The efficiency and effectiveness of the editorial review process is critically dependent upon the actions of both authors submitting papers and the reviewers. Authors accept the responsibility of preparing research papers at a level suitable for evaluation by independent reviewers. Such preparation, therefore, should include subjecting the manuscript to critique by colleagues and others and revising it accordingly prior to submission. The review process is not to be used as a means of obtaining feedback at early stages of developing the research.

Reviewers and associate editors are responsible for providing critically constructive and prompt evaluations of submitted research papers based on the significance of their contribution and on the rigor of analysis and presentation. Associate editors also make editorial recommendations to the editor.

MANUSCRIPT PREPARATION AND STYLE

The Accounting Review's manuscript preparation guidelines follow (with a slight modification) the B-format of *The Chicago Manual of Style* (14th ed.; University of Chicago Press). Another helpful guide to usage and style is *The Elements of Style*, by William Strunk, Jr., and E. B. White (Macmillan). Spelling follows *Webster's International Dictionary*.

FORMAT

1. All manuscripts should be typed in 12-point font on one side of 8½ × 11" good quality paper and be double spaced, except for indented quotations.
2. Manuscripts should be as concise as the subject and research method permit, generally not to exceed 7,000 words.
3. Margins of at least one inch from top, bottom, and sides should facilitate editing and duplication.
4. To promote anonymous review, authors should not identify themselves directly or indirectly in their papers or in experimental test instruments included with the submission. Single authors should not use the editorial "we."
5. A cover page should show the title of the paper, the author's name, title and affiliation, email address, any acknowledgments, and a footnote indicating whether the author would be willing to share the data (see last paragraph in this statement).

Pagination: All pages, including tables, appendices and references, should be serially numbered. Major sections should be numbered in Roman numerals. Subsections should not be numbered.

Numbers: Spell out numbers from one to ten, except when used in tables and lists, and when used with mathematical, statistical, scientific, or technical units and quantities, such as distances, weights and measures. For example: *three days*; *3 kilometers*; *30 years*. All other numbers are expressed numerically.

Percentages and Decimal Fractions: In nontechnical copy use the word percent in the text.

Hyphens: Use a hyphen to join unit modifiers or to clarify usage. For example: *a well-presented analysis*; *re-form*. See *Webster's* for correct usage.

Keywords: The abstract is to be followed by four keywords that will assist in indexing the paper.

ABSTRACT/INTRODUCTION

An Abstract of about 100 words should be presented on a separate page immediately preceding the text. The Abstract should concisely inform the reader of the manuscript's topic, its methods, and its findings.

Keywords and the Data Availability statements should follow the Abstract. The text of the paper should start with a section labeled "I. Introduction," which provides more details about the paper's purpose, motivation, methodology, and findings. Both the Abstract and the introduction should be relatively nontechnical, yet clear enough for an informed reader to understand the manuscript's contribution. The manuscript's title, but neither the author's name nor other identification designations, should appear on the Abstract page.

TABLES AND FIGURES

The author should note the following general requirements:

1. Each table and figure (graphic) should appear on a separate page and should be placed at the end of the text. Each should bear an Arabic number and a complete title indicating the exact contents of the table or figure. Tables and figures should define each variable. The titles and definitions should be sufficiently detailed to enable the reader to interpret the tables and figures without reference to the text.
2. A reference to each graphic should be made in the text.
3. The author should indicate by marginal notation where each graphic should be inserted in the text.
4. Graphics should be reasonably interpreted without reference to the text.
5. Source lines and notes should be included as necessary.

Equations: Equations should be numbered in parentheses flush with the right-hand margin.

DOCUMENTATION

Citations: Work cited should use the "author-date system" keyed to a list of works in the reference list (see below). Authors should make an effort to include the relevant page numbers in the cited works.

1. In the text, works are cited as follows: authors' last name and date, without comma, in parentheses: for example, (Jones 1987); with two authors: (Jones and Freeman 1973); with more than two: (Jones et al. 1985); with more than one source cited together (Jones 1987; Freeman 1986); with two or more works by one author: (Jones 1985, 1987).
2. Unless confusion would result, do not use "p." or "pp." before page numbers: for example, (Jones 1987, 115).
3. When the reference list contains more than one work of an author published in the same year, the suffix a, b, etc. follows the date in the text citation: for example, (Jones 1987a) or (Jones 1987a; Freeman 1985b).
4. If an author's name is mentioned in the text, it need not be repeated in the citation; for example, "Jones (1987, 115) says...."
5. Citations to institutional works should use acronyms or short titles where practicable; for example, (AAA ASOBAT 1966); (AICPA Cohen Commission Report 1977). Where brief, the full title of an institutional work might be shown in a citation: for example, (ICAEW *The Corporate Report* 1975).
6. If the manuscript refers to statutes, legal treatises or court cases, citations acceptable in law reviews should be used.

Reference List: Every manuscript must include a list of references containing only those works cited. Each entry should contain all data necessary for unambiguous identification. With the author-date system, use the following format recommended by *The Chicago Manual*:

1. Arrange citations in alphabetical order according to surname of the first author or the name of the institution responsible for the citation.
2. Use author's initials instead of proper names.
3. Date of publication should be placed immediately after author's name.
4. Titles of journals should not be abbreviated.
5. Multiple works by the same author(s) in the same year are distinguished by letters after the date.
6. Inclusive page numbers are treated as recommended in *Chicago Manual* section 8.67.

Sample entries are as follows:

- American Accounting Association, Committee on Concepts and Standards for External Financial Reports. 1977. *Statement on Accounting Theory and Theory Acceptance*. Sarasota, FL: AAA.
- Demski, J. S., and D. E. M. Sappington. 1989. Hierarchical structure and responsibility accounting. *Journal of Accounting Research* 27 (Spring): 40–58.
- Dye, R., B. Balachandran, and R. Magee. 1989. Contingent fees for audit firms. Working paper, Northwestern University, Evanston, IL.

- Fabozzi, F., and I. Pollack, eds. 1987. *The Handbook of Fixed Income Securities*. 2nd edition. Homewood, IL: Dow Jones-Irwin.
- Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, U.K.: Cambridge University Press.
- Porcano, T. M. 1984a. Distributive justice and tax policy. *The Accounting Review* 59 (October): 619–636.
- . 1984b. The perceived effects of tax policy on corporate investment intentions. *The Journal of the American Taxation Association* 6 (Fall): 7–19.
- Shaw, W. H. 1985. Empirical evidence on the market impact of the safe harbor leasing law. Ph.D. dissertation, The University of Texas at Austin.
- Sherman, T. M., ed. 1984. *Conceptual Framework for Financial Accounting*. Cambridge, MA: Harvard Business School.

Footnotes: Footnotes are not used for documentation. Textual footnotes should be used only for extensions and useful excursions of information that if included in the body of the text might disrupt its continuity. Footnotes should be consecutively numbered throughout the manuscript with superscript Arabic numerals. Footnote text should be doubled-spaced and placed at the end of the article.

SUBMISSION OF MANUSCRIPTS

Authors should note the following guidelines for submitting manuscripts:

1. Manuscripts currently under consideration by another journal or publisher should not be submitted. The author must state that the work is not submitted or published elsewhere.
2. In the case of manuscripts reporting on field surveys or experiments, four copies of the instrument (questionnaire, case, interview plan or the like) should be submitted. Information that might identify the author(s) must be deleted from the instrument.
3. Four copies should be submitted together with a check in U.S. funds for \$75.00 for members or \$100.00 for nonmembers of the AAA made payable to the American Accounting Association. Effective January 1990, the submission fee is nonrefundable.
4. The author should retain a copy of the paper.
5. Revisions must be submitted within 12 months from request, otherwise they will be considered new submissions.

COMMENTS

Comments on articles previously published in *The Accounting Review* will be reviewed (anonymously) by two reviewers in sequence. The first reviewer will be the author of the original article being subjected to critique. If substance permits, a suitably revised comment will be sent to a second reviewer to determine its publishability in *The Accounting Review*. If a comment is accepted for publication, the original author will be invited to reply. All other editorial requirements, as enumerated above, also apply to proposed comments.

POLICY ON REPRODUCTION

An objective of *The Accounting Review* is to promote the wide dissemination of the results of systematic scholarly inquiries into the broad field of accounting.

Permission is hereby granted to reproduce any of the contents of the *Review* for use in courses of instruction, as long as the source and American Accounting Association copyright are indicated in any such reproductions.

Written application must be made to the American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399, for permission to reproduce any of the contents of the *Review* for use in other than courses of instruction—e.g., inclusion in books of readings or in any other publications intended for general distribution. In consideration for the grant of permission by the *Review* in such instances, the applicant must notify the author(s) in writing of the intended use to be made of each reproduction. Normally, the *Review* will not assess a charge for the waiver of copyright.

Except where otherwise noted in articles, the copyright interest has been transferred to the American Accounting Association. Where the author(s) has (have) not transferred the copyright to the Association, applicants must seek permission to reproduce (for all purposes) directly from the author(s).

POLICY ON DATA AVAILABILITY

The following policy has been adopted by the Executive Committee in its April 1989 meeting.

"An objective of (*The Accounting Review, Accounting Horizons, Issues in Accounting Education*) is to provide the widest possible dissemination of knowledge based on systematic scholarly inquiries into accounting as a field of professional research, and educational activity. As part of this process, authors are encouraged to make their data available for use by others in extending or replicating results reported in their articles. Authors of articles which report data dependent results should footnote the status of data availability and, when pertinent, this should be accompanied by information on how the data may be obtained."

APRIL 2002 PLACEMENT ADS

The deadline for information to be included in the "Placement Information" section is two months prior to the date of publication. The Review is published on the first day of January, April, July, and October. Placement advertising and replies should be sent to the American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399, ATTENTION: ADVERTISING DEPT. All replies to box numbers should be on separate sheets to facilitate forwarding.

- 201 MICHIGAN STATE UNIVERSITY, Eli Broad Graduate School of Management, seeks to hire one tenure-track Assistant Professor faculty member in Financial Accounting beginning August 2002. Michigan State provides an excellent environment for new faculty—competitive salary and benefits, summer research support, and an attractive teaching load. An earned doctorate is required and professional certification and teaching experience in Financial Accounting are desirable. ABD candidates in advanced stages of the dissertation will be considered. Review of applications will begin November 26, and continue until position is filled. Candidates are asked to submit an application letter, resume, and the names, addresses, and phone numbers of three references. Contact Professor Thomas J. Linsmeier, N256 Business College Complex, Eli Broad College of Business, Michigan State University, East Lansing, MI 48824-1121; Email: tjl@msu.edu. Michigan State University is an Affirmative Action/Equal Opportunity Institution. Handicappers have the right to request and receive reasonable accommodations.
- 202 UNIVERSITY OF WISCONSIN-PLATTEVILLE invites applications for an Assistant Professor, nine-month, tenure-track position available August 24, 2002. Earned doctorate in Accounting and CPA with specialization in Auditing is preferred. M.B.A. or master's degree in Accounting with specialization in Auditing, Systems, or Taxation plus CPA or other appropriate designation; J.D. with Accounting background; or qualifications comparable to either of the above, required. Excellent verbal and written communication skills, willingness to teach students at a distance using alternative delivery technologies, and demonstrable commitment to a diverse learning environment are required. At least two years of college teaching in Accounting, two years of practical experience, and experience teaching students at a distance preferred. Send letter of application, resume, transcripts, and three letters of recommendation to Jack Borke, Search Committee, Dept. of Business and Accounting, UW-Platteville, One University Plaza, Platteville, WI 53818. Web site: <http://www.uwplatt.edu/~pers/employ.htm>. Review of applications begins November 26, 2001 and will continue until the position is filled.
- 203 WAYNE STATE UNIVERSITY, School of Business Administration, Department of Accounting, is seeking full-time, tenure-track faculty for the 2002–2003 academic year. Qualifications for the position include an earned doctoral degree in Accounting, with primary focus on Financial Accounting and publications consistent with rank, and a demonstrated commitment to quality teaching. The School of Business Administration, accredited by AACSB International, enrolls 1,000 undergraduate students and 1,800 graduate students. The Accounting Department offers baccalaureate and master's degrees. Salary and benefit package is highly competitive. Review of applications will begin immediately and continue until the positions are filled. Please send a letter, *curriculum vitae*, and the names of three references to Assistant Dean of Administrative Affairs, 105 Prentiss Building, School of Business Administration, Wayne State University, Detroit, MI 48202. Committed to diversity, the school actively encourages applications from women and minority persons. WSU is an Equal Opportunity/Affirmative Action Employer.

- 204 KDI School of Public Policy and Management, Seoul, Korea, invites applications for a tenure-track position starting Fall 2002. Applicants interested in a visiting position are also welcome. Ph.D. or D.B.A. in Accounting is required, and CPA is preferred. Candidates at the dissertation stage are also encouraged to apply. KDI School offers internationally competitive salary, research support, reduced teaching loads for research-active faculty, funding to attend international conferences, and relocation assistance and benefits to help international faculty take up their appointments in Seoul, Korea. Detailed information about the KDI School and application procedures are available at the KDI web site: http://www.kdischool.ac.kr/whats/posi_200112.htm. Send letter of application, resume, letter of recommendation, and recent research papers to Ms. Soo-Kyoung Chung, Office of Planning Affairs, KDI School of Public Policy and Management, Cheongnyangri-dong, Dongdaemun-gu, Seoul, Korea 130-868; Email: kyoung@kdischool.ac.kr. Review of applications will continue until the position is filled.
- 205 SONOMA STATE UNIVERSITY invites applications for a tenure-track Accounting position at the Assistant to Full Professor level for Fall 2002. Applicants must have a doctoral degree in Accounting and have demonstrated potential for competence in teaching, research, and service. Preference will be given to those candidates holding a CPA license. Teaching and professional experience are highly desirable. Applicant screening will begin on February 1, 2002 and will continue until the position is filled. Please submit a letter of application, a *curriculum vitae*, current teaching evaluations, if available, and the names, titles, and telephone numbers of three references to Chair, Accounting Search Committee, Department of Business Administration, Sonoma State University, 1801 E. Cotati Avenue, Rohnert Park, CA 94928-3609. SSU is part of the California State University System and is located 48 miles north of San Francisco.
- 206 KING'S COLLEGE in Wilkes-Barre, PA is seeking an Assistant or Associate Professor in Accounting starting Fall 2002. A completed Ph.D. or D.B.A. is preferred; an ABD with established completion date will be considered. Salary will be competitive with northeastern liberal arts institutions. Rank will depend on experience. King's is an independent co-educational institution, enrolling 1,800 full-time and 450 part-time undergraduate students and 100 graduate students. In an open Catholic tradition, King's fosters religious, moral, and social development. A rigorous curriculum of liberal studies provides a foundation for nearly 30 major programs. King's is nationally recognized for its innovations in liberal arts education. Candidates should send a letter outlining teaching and research interests, *curriculum vitae*, transcripts, and three letters of recommendation to David G. Martin, Ph.D., CPIM, Director, McGowan School of Business, King's College, Wilkes-Barre, PA 18711. King's College is an Equal Opportunity Employer and specifically invites and encourages applications from women and minorities.
- 207 UNIVERSITY OF GEORGIA, J.M. Tull School of Accounting, is seeking applications for an anticipated tenure-track position in Accounting at the Assistant Professor level with teaching responsibilities in Taxation or Systems/Auditing at both the graduate and undergraduate levels. The anticipated date of appointment is Fall 2002. Candidates for this position should have either a Ph.D. in Accounting or be well along in the dissertation stage. Applicants must have a commitment to excellence in teaching and scholarship. Salary and benefits are competitive. Applications received before February 28, 2002 are assured of consideration. Please submit application materials (current *curriculum vitae* and dissertation proposal or example of scholarly work) to Professor Ben Ayers, Chair, Faculty Recruiting Committee, J.M. Tull School of Accounting, Terry College of Business, 255 Brooks Hall, University of Georgia, Athens, GA 30602. The University of Georgia is an Equal Opportunity/Affirmative Action Institution and strongly encourages applications from women and minorities.

- 208 UNIVERSITY OF CALIFORNIA, Los Angeles invites applications for positions in Accounting at both the junior and senior levels beginning July 1, 2002. Applicants must have a Ph.D. degree or be a Ph.D. candidate and exhibit outstanding research capabilities. *Curriculum vitae* and working papers should be sent to Professor Carla Hayn, The John E. Anderson Graduate School of Management, Box 951481, UCLA, Los Angeles, CA 90095-1481. Under federal law, the University of California may employ only individuals who are legally authorized to work in the United States as established by providing documents specified in the Immigration Reform and Control Act of 1986. UCLA is an Affirmative Action/Equal Opportunity Employer.
- 209 WIDENER UNIVERSITY, Department of Accounting and Taxation, invites applications for a tenure-track position at the rank of Assistant/Associate Professor, starting Fall 2002. Candidates' primary teaching interests should be in Auditing/Managerial Accounting. Interest in Accounting Information Systems is desirable. A Ph.D. or D.B.A., completed or in advanced stages, is required; professional certification a plus. Candidates must have demonstrated potential for excellence in teaching and research, and will contribute in further developing an information-technology-focused curriculum. The SBA, which is AACSB-accredited, offers a major in Accounting, an M.S. in Accounting Information Systems, M.S. in Taxation, and M.B.A. Interviews will be held at the Mid-Atlantic AAA Meeting in Baltimore, MD on April 26, 2002 by advance appointment only. Contact and send *curriculum vitae* to Dr. Joseph Hargadon, Widener University, One University Place, Chester, PA 19013; Phone: (610) 499-4280; Email: joseph.m.hargadon@widener.edu.
- 210 KENT STATE UNIVERSITY, the Department of Accounting, invites applications for an Assistant Professor with teaching and research emphasis in Financial, Managerial, and/or Information Systems beginning in Fall 2002. A Ph.D. with a major in Accounting or an appropriate related field is required; professional certification such as CPA or CMA is highly desirable. It is preferred that applicants have their doctorate completed or expect to have all degree requirements completed by beginning date. Candidates who are ABD will be considered. Evidence of the potential for effective undergraduate and graduate teaching, and discovery/integration research is required. Research will be expected in high-quality, academic refereed journals for reappointment, tenure, and promotion; some publication in other journals may be included in the faculty member's portfolio. Participation in teaching doctoral students and serving on and/or supervising their dissertations likely will be part of this faculty member's responsibilities. If you are interested in this position, please contact and/or mail a copy of your resume to Dr. Pervaiz Alam, Chair, Faculty Search Committee, Department of Accounting, 576 BSA, Kent State University, Kent, OH 44242; Email: palam@bsa3.kent.edu; Phone (330) 672-1121; Fax (330) 672-2548.
- 211 LOYOLA COLLEGE IN MARYLAND invites applications for a tenure-track position at the Assistant or Associate level in Financial Accounting. Candidates for the position must possess an appropriate earned doctorate or be near completion and have a commitment to outstanding teaching, quality research, and dedicated service. The primary teaching area is Financial Accounting, with Accounting Information Systems as a possible secondary area. Appropriate business and teaching experience are highly valued. The Sellinger School of Business and Management, and the Department of Accounting and Information Systems are AACSB-accredited. Competitive salary, attractive benefits, and many technology and financial supports are available. For more information about the position, visit our web site at <http://www.loyola.edu>. Submit application to Dr. Alfred Michenzi, Department of Accounting and Information Systems, Loyola College in Maryland, 4501 North Charles Street, Baltimore, MD 21210-2699, Email: amichenzi@loyola.edu. Loyola College is a Jesuit Catholic institution that welcomes applicants from all backgrounds who can contribute to our unique educational mission. Loyola is an Equal Opportunity Employer that seeks applications from women and members of minority groups.